



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 140719

TO: James Schultz
Location: REM-2D18/2C18
Art Unit: 1635
Wednesday, December 22, 2004

Case Serial Number: 10/001851

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Schultz,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

no page blank (uspto)

STIC-Biotech/ChemLib

140719

From: Schultz, James
Sent: Sunday, December 19, 2004 9:56 PM
To: STIC-Biotech/ChemLib
Subject: Seq search 10/001,851

Hello,
Could you please search SEQ ID NO 2 in the amino acid databases (603 aa long, no interference search needed) in the above entitled application?
Thanks
Doug Schultz

James Douglas Schultz, PhD
AU 1635 (Biotechnology)
Patent Examiner
United States Patent and Trademark Office
(Office) REM 2D18
(Mail) REM 2C18
(571) 272-0763

RECEIVED
DEC 20 2005
STIC

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: 12/20/04
Date Completed: _____
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search
NA Sequence: # _____
AA Sequence: # 1
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

Image Blank (uspto)

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions .rnpm and .rnpn

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions .rapm and .rapn

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

Page Blank (uspto)

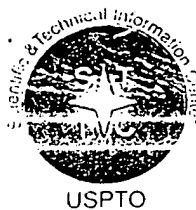
Protein Sequence Searches - 10/8/04

All of the sequence databases on the ABSS have been updated. A change has occurred in the protein databases.

- Two protein databases, SPTREMBL and SwissProt, are now produced as a single, merged database called UniProt.
- Results from UniProt have the file extension **.rup**.
- Sequences in UniProt are identified by the same ID that had been used in SPTREMBL or SwissProt.
- In instances where the database curators have determined that an SPTREMBL record and a SwissProt record represent the same sequence, the two records have been merged into one. Both IDs are present in the record. Any differences found between the two sequences are recorded in the FT (feature table) fields.

If you have any questions regarding these changes or your results, please contact any STIC searcher.

nis Page Blank (uspto)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
571-272-2507 Remsen E01 D86

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library Remsen Bldg.



This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      December 20, 2004, 13:55:08 ; Search time 159 Seconds
              (without alignments)
              1360.464 Million cell updates/sec
```

Title: US-10-001-851-2
Perfect score: 3278
Sequence: 1 MRKREKRLQAVLVLAIV.....TQQWLFHTNSTVLEKENRN 6030

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues
Total number of hits satisfying chosen parameters: 2002273

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
1: geneexpq123sep04:*
2: geneexpq1980s:*
3: geneexpq1990s:*
4: geneexpq2000s:*
5: geneexpq2001s:*
6: geneexpq2002s:*
7: geneexpq2003as:*
8: geneexpq2003bs:*
9: geneexpq2004as:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3278	100.0	603	4	AAU07778	Aau07778 Human nov
2	3278	100.0	603	5	AAU99098	Aau99098 Human gly
3	3272	99.8	603	7	ADCS3387	Adcs3387 Human N-a
4	3242.5	98.9	631	4	AAU07777	Aau07777 Human nov
5	2771	84.5	506	4	AAU07772	Aau07772 Human nov
6	2771	84.5	506	6	AAG79781	Aag79781 Carbohydr
7	2746.5	83.8	535	4	AAU07771	Aau07771 Human nov
8	1893	57.7	339	4	AAB94733	Aab94733 Human prot
9	1877	57.3	366	4	AAU07776	Aau07776 Human nov
10	1818	55.5	407	4	AAG62600	Aag62600 Human ace
11	1403.5	42.8	319	4	ABE08187	Abeg08187 Novel hum
12	1396	42.6	275	7	ABE25865	Abeg25865 Human N-a
13	1392	42.5	276	8	AD011613	Adq11613 Human soB
14	1385.5	42.3	335	4	ABE08189	Abg08189 Novel hum
15	1370	41.8	269	4	AAU07770	Aau07770 Human nov
16	1360	41.5	666	4	ABB66873	Abb66873 Drosophil
17	1360	41.5	666	4	ABB58867	Abb58867 Drosophil
18	1311	40.0	321	4	AAU07774	Aau07774 Human nov
19	1186	36.2	1252	4	ABBS5363	Abb55363 Drosophil
20	1155	35.2	499	4	ABG08190	Abg08190 Novel hum
21	1125	34.3	559	7	ADP65304	Adp65304 Human can
22	1125	34.3	561	3	AAB43561	Aab43561 Human can
23	1116	34.0	559	7	ADB79836	Adb79836 Rat polyP
24	1116	34.0	559	7	ADB79796	Adb79796 Rat polyP
25	1116	34.0	559	7	ADB62526	Adb62526 Rat Prote

26	1115	517	2	AAR664402	Aar664402	GalNac-1T
27	1115	517	2	AAM164889	Aam164889	Honeybee
28	1115	559	2	AAR66397	Aar66397	Cattle Ga
29	1115	559	2	AAR66401	Aar66401	GalNac-1T
30	1115	559	2	AAM164884	Aaw164884	Bovine N-
31	1113.5	630	4	ABB64271	Abb64271	Drosophil
32	1112	571	7	ADBC10116	Adbc10116	Human NOV
33	1106	556	5	ABG32379	Abg32379	Novel hum
34	1106	556	7	ADB25818	Adb25818	Novel N-a
35	1106	556	7	ADB76895	Adb76895	Human N-a
36	1096.5	555	7	ADBC10114	Adbc10114	Human NOV
37	1085.5	573	8	ADP04927	Adp04927	Sea squit
38	1050	657	4	AAR88419	Aab88419	Human mem
39	1048	478	6	ABR45118	Abri45118	Human diti
40	1048	478	7	ADJ22636	Adj22636	Human dis
41	1045.5	658	4	AAR78711	Aam78711	Human pro
42	1040	580	7	ADBS5070	Adbs5070	Rat prote
43	1035.5	591	4	ABB62119	Abb62119	Drosophil
44	1026	578	6	ABB07527	Abb07527	Human N-a
45	1024.5	188	7	ADB25974	Adb25974	N-acetylgl

ALIGNMENTS

```

RESULT 1
AAU07778
ID   AAU07778 standard; protein; 603 AA.

```

AC MAU07778;
VY

DT 04-DEC-2001 (first entry)

DE Human novel transferase protein, NHP #21.

KW Human; transferase; breast cancer; prostate cancer; immunogen;

KM gene therapy; antisense.

OS Homo sapiens.

PN WO200164903-A2.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001WO-US006460.

PR 29-FEB-2000; 2000US-0185920P.

PR 24-MAR-2000; 2000US-0191849P.

PA (LEXI-) LEXICON GENETICS INC.

PA (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;
CT Zamboni P, Conda M, Wallis DW, Wilczewski W, Hu Y, Wicko TA.

PI Potter DG;

WPI; 2001-550185/61

XXXX

PT diagnosis, drug screening, clinical trial monitoring and treatment of

XX

XX

CC transferase proteins (N

CC diseases and disorders

CC useful for augmenting the efficacy of chemotherapeutic agents used in
CC treatment of breast or prostate cancer. The nucleic acid is also useful
CC in NHP gene regulation, and as antisense primers in amplification
CC reactions of NHP gene sequences. NHPs are useful for producing

CC antibodies. The present sequence represents a novel human transferase
CC which has sequence similarity to N-acetyl-galactosaminyltransferase
XX
SQ Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRREKRLIOAVLVLAALVLPVNGMLTYREROPDGTGGGAAPAAAGGSHSRK 60
DB 1 MRREKRLIOAVLVLAALVLPVNGMLTYREROPDGTGGGAAPAAAGGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDARVNGEGRPYPMTDARVDAVRENGENIYVSDK 120
DB 61 KTFPLDGGQKLKDMHDKAIRDARVNGEGRPYPMTDARVDAVRENGENIYVSDK 120
QY 121 ISLNSRLPIRHPNCNSKRYLETLPNTSIIIPPHNEGSSLLRTVHSVLRSPPELVAEI 180
DB 121 ISLNSRLPIRHPNCNSKRYLETLPNTSIIIPPHNEGSSLLRTVHSVLRSPPELVAEI 180
QY 181 VLVDDSDREHLKKPLEDMALFPSVRIITKKREGILRTRMGASVATGDTTFLDSHC 240
DB 181 VLVDDSDREHLKKPLEDMALFPSVRIITKKREGILRTRMGASVATGDTTFLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIP 300
DB 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIP 300
QY 301 PELQKADPSDPFSPVMAAGLPAVDRKFMWELGYPDGLIWEQGEYIEISFKVMCGGRM 360
DB 301 PELQKADPSDPFSPVMAAGLPAVDRKFMWELGYPDGLIWEQGEYIEISFKVMCGGRM 360
QY 361 EDIPGSRVGHITKRYVPYKIPAGVSLARNLKRVAEWMMDYATYTORREYHLSAGDV 420
DB 361 EDIPGSRVGHITKRYVPYKIPAGVSLARNLKRVAEWMMDYATYTORREYHLSAGDV 420
QY 421 AVOKLRSSLNCSFKPMFTKIAMDLPKFPVPEPPAAAGELIRNVGTGCACTKIGALG 480
DB 421 AVOKLRSSLNCSFKPMFTKIAMDLPKFPVPEPPAAAGELIRNVGTGCACTKIGALG 480
QY 481 SPRLTSGCVARGGEAAMNNMNVFTFTWRREDIRGDDPQHTKKCFDASHSTSVLYDCHS 540
DB 481 SPRLTSGCVARGGEAAMNNMNVFTFTWRREDIRGDDPQHTKKCFDASHSTSVLYDCHS 540
QY 541 MKGNOLMKYRKDKTIVHPVSGCMDCSESDHRIFMNTCNSSLTQQLFEHTNSTYLEKF 600
DB 541 MKGNOLMKYRKDKTIVHPVSGCMDCSESDHRIFMNTCNSSLTQQLFEHTNSTYLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 2
AAU99098
ID AAU99098 standard; protein; 603 AA.
XX
AC AAU99098;

30-AUG-2002 (first entry)
XX
DE Human glycosyl transferase 47169.
XX

KM Human; glycosyl transferase; 47169; 33935; cancer; carbohydrate storage;
KM diabetes mellitus; hypoglycaemia; arthritis; rheumatism;
KM autoimmune disorder; systemic lupus erythematosus; Grave's disease;
KM myasthenia gravis; insulin resistance; scleroderma; rheumatoid arthritis;
KM autoimmune infertility; tumourigenesis.

OS Homo sapiens.
XX
XX
PN W0200240657-A2.

XX 23-MAY-2002.
PD
XX
PF 20-NOV-2001; 2001WO-US047575.
PR
XX
PR 20-NOV-2000; 2000US-0249939P.
PA
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Meyers R, Williamson M;
XX
PI WPI; 2002-508326/54.
DR
XX
PS N-PSDB; ABK86093.

Use of modulator of activity of novel glycosyl transferase proteins,
47169/33935 proteins, for making a medicament for modulating ability of
cell to affect glycosylation state of lipid or polypeptide target in
cell.
Claim 27; Fig 1; 153pp; English.

The invention relates to the use of a modulator of the activity of an
enzyme such as 47169 or 33935 protein (a novel glycosyl transferase
or protein) for making a medicament for modulating the ability of a cell to
affect the glycosylation state of a lipid target or polypeptide target in
a cell. Also included is a method of assessing (M1) if a test compound is
useful for modulating at least one phenomenon (P) such as non-covalent
binding between a protein and one of a cell, a virus and another protein;
cell signaling; cell differentiation; tumourigenesis; cell adhesion; cell
motility; cell-to-cell interaction; cell invasivity; cell proliferation;
gene transcription; and an immune response, comprising: (a) adding the
test compound to a first composition comprising a 603 residue 47169
polypeptide sequence (S2), or a sequence at least 90 % identical to a 492
residue 33935 polypeptide sequence (S12) both given in the specification
, and (b) comparing the activity in the first composition and in a second
composition that is substantially identical to the first composition,
except that it lacks the test compound, whereby a difference in the
activity in the first and second compositions is an indication that the
test compound is useful for modulating the phenomenon. The method is
useful for making a medicament for modulating the ability of a cell (e.g.
human endothelial cell such as lung cell, breast cell or colon cell,
preferably a tumour cell) to affect the glycosylation state of a target
such as lipid or polypeptide. The identified test compound is useful for
treating disorders such as diabetes mellitus, hypoglycaemia, arthritis,
rheumatism, autoimmune disorders (e.g. systemic lupus erythematosus,
Grave's disease, myasthenia gravis, insulin resistance, rheumatoid
arthritis, scleroderma and autoimmune infertility), tumourigenesis,
cancer and tumour metastasis. The present sequence represents glycosyl
transferase 47169

SQ Sequence 603 AA;

Query Match 100.0%; Score 3278; DB 5; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRREKRLIOAVLVLAALVLPVNGMLTYREROPDGTGGGAAPAAAGGSHSRK 60
DB 1 MRREKRLIOAVLVLAALVLPVNGMLTYREROPDGTGGGAAPAAAGGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDARVNGEGRPYPMTDARVDAVRENGENIYVSDK 120
DB 61 KTFPLDGGQKLKDMHDKAIRDARVNGEGRPYPMTDARVDAVRENGENIYVSDK 120
QY 121 ISLNSRLPIRHPNCNSKRYLETLPNTSIIIPPHNEGSSLLRTVHSVLRSPPELVAEI 180
DB 121 ISLNSRLPIRHPNCNSKRYLETLPNTSIIIPPHNEGSSLLRTVHSVLRSPPELVAEI 180
QY 181 VLVDDSDREHLKKPLEDMALFPSVRIITKKREGILRTRMGASVATGDTTFLDSHC 240
DB 181 VLVDDSDREHLKKPLEDMALFPSVRIITKKREGILRTRMGASVATGDTTFLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYTKRIP 300

```

Db      241 EAVVNLPLLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Qy      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPGLEIWCGEYBISFKYMMCGGRM 360
Db      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPGLEIWCGEYBISFKYMMCGGRM 360
Qy      361 EDIPGSRVGHYIKKYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Db      361 EDIPGSRVGHYIKKYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Qy      421 AVQKLRSSLNCKSPFKMFTKIAMDLPKFPYVPEPPAAWGEIRNVGTGLCADTKIGALG 480
Db      421 AVQKLRSSLNCKSPFKMFTKIAMDLPKFPYVPEPPAAWGEIRNVGTGLCADTKIGALG 480
Qy      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Db      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Qy      541 MKGNOLMKYRKDKTLVHPVSGSCMDSESDBRIFMNTCNPSLSLTOQWLFHTNISTYLEKF 600
Db      541 MKGNOLMKYRKDKTLVHPVSGSCMDSESDBRIFMNTCNPSLSLTOQWLFHTNISTYLEKF 600
Qy      601 NRN 603
Db      601 NRN 603

RESULT 3
ADC35387
ID ADC35387 standard; protein; 603 AA.
XX
XX ADC35387;
XX
XX 18-DEC-2003 (first entry)
XX
DE Human N-acetyl-galactosamine transferase (GalNac) T13 protein.
XX
XX N-acetyl-galactosamine transferase; GalNac; alpha1-bond;
XX genetic engineering; GalNac-T13; cyostatic; cancer treatment;
XX cancer diagnosis; gene therapy; human; enzyme.
XX
XX Homo sapiens.
XX
XX MO2003057887-A1.
XX
XX 17-JUL-2003.
XX
XX 06-JAN-2003; 2003WO-JP000008.
XX
XX 28-DEC-2001; 2001JP-00401507.
XX 05-JUN-2002; 2002JP-00163832.
XX 12-JUL-2002; 2002JP-00203696.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (AMSH) AMERSHAM BIOSCIENCES KK.
XX (FURE) FUJIREBIO INC.
XX
XX Narimatsu H, Zhang Y, Gotoh M;
XX
XX WPI; 2003-587133/55.
XX DR N-PSDB; ADC35394.
XX
XX Novel UDP-N-acetyl-D-galactosamine:polypeptide N-acetyl-galactosamine
XX transferase and encoded nucleic acid, applicable in identifying O-bonded
XX sugar-attached proteins, gene diagnosis, and cancer therapy.
XX
XX Claim 1; SEQ ID NO 3; 153pp; Japanese.
XX
XX This invention relates to a novel protein with an activity of
XX transferring N-acetyl-galactosamine (GalNac) via an alpha1-bond to a
XX hydroxyl group of serine or threonine in a protein or a peptide sequence.
XX The invention provides gene and protein sequences for these enzymes to

```

```

CC enable genetic engineering or production of these enzymes. Seven genes
CC and their proteins are described by the invention, GalNac-T11 to 17. The
CC enzymes of the invention may have cyostatic activity, and hence may be
CC used for treatment or diagnosis of cancer, and the sequences described
CC may be useful in gene therapy. The present sequence is the partial
CC sequence of the human GalNac-T13 transferase protein of the invention.
XX
XX Sequence 603 AA:
Qy      1 MRKREKRLQAVLVAALVLPVNGVLMVLYRROPDGRPGSGAAVAPAGGSGHSRQK 60
Db      1 MRKREKRLQAVLVAALVLPVNGVLMVLYRROPDGRPGSGAAVAPAGGSGHSRQK 60
Qy      61 KTFPLDGGOKLQDMHDKAIRDQORVNGEQGRPYPMTDABRVDAVRENGFNIVYSDK 120
Db      61 KTFPLDGGOKLQDMHDKAIRDQORVNGEQGRPYPMTDABRVDAVRENGFNIVYSDK 120
Qy      121 ISLNRSLPDIRHPNCSKYLETLPTNTSIIIPPHNGWSLRTVHSVLRSPPELVAEI 180
Db      121 ISLNRSLPDIRHPNCSKYLETLPTNTSIIIPPHNGWSLRTVHSVLRSPPELVAEI 180
Qy      181 VLVDDFSDEHLLKKPLLEDVVALPSPVRLIRTKRREGILIRMLGASVATGDTTFLDSHC 240
Db      181 VLVDDFSDEHLLKKPLLEDVVALPSPVRLIRTKRREGILIRMLGASVATGDTTFLDSHC 240
Qy      241 EAVVNLPLLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Db      241 EAVVNLPLLDRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYKRIPIR 300
Qy      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPGLEIWCGEYBISFKYMMCGGRM 360
Db      301 PELQKADPSDPFSPVWAGGLFVDRKMFWEELGYPGLEIWCGEYBISFKYMMCGGRM 360
Qy      361 EDIPGSRVGHYIKKYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Db      361 EDIPGSRVGHYIKKYVPYKVPAGVSLARNLKRVAEVMDEYAEYIYORREYHLSAGDV 420
Qy      421 AVQKLRSSLNCKSPFKMFTKIAMDLPKFPYVPEPPAAWGEIRNVGTGLCADTKIGALG 480
Db      421 AVQKLRSSLNCKSPFKMFTKIAMDLPKFPYVPEPPAAWGEIRNVGTGLCADTKIGALG 480
Qy      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Db      481 SPLRLGCVRGGEAAMNNMNVFTFTWRREDIRPGDPQHTKKFCFDAISHTSPVTLVDCHS 540
Qy      541 MKGNOLMKYRKDKTLVHPVSGSCMDSESDBRIFMNTCNPSLSLTOQWLFHTNISTYLEKF 600
Db      541 MKGNOLMKYRKDKTLVHPVSGSCMDSESDBRIFMNTCNPSLSLTOQWLFHTNISTYLEKF 600
Qy      601 NRN 603
Db      601 NRN 603

RESULT 4
AAU07777
ID AAU07777 standard; protein; 631 AA.
XX
XX AAU07777;
XX
XX 04-DEC-2001 (first entry)
XX
DE Human novel transferase protein, NHP #20.
XX
XX Human; transferase; breast cancer; prostate cancer; immunogen;
XX gene therapy; antisense.
XX
XX Homo sapiens.
XX

```

Key Location/Qualifiers
 MISC-difference 630..631
 /note="Encoded by AATGGAAC"

XX Key
 XX MISC-difference 630..631
 XX /note="Encoded by AATGGAAC"

XX WO200164903-A2.
 XX 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US006460.
 XX 29-FEB-2000; 2000US-0185920P.
 XX 02-MAR-2000; 2000US-0186558P.
 XX 24-MAR-2000; 2000US-0191849P.
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A,
 XX Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
 XX Potter DG;
 XX MPI; 2001-550185/61.
 XX N-PSDB; AAS12624.

XX Novel nucleic acid sequences encoding novel human proteins useful for
 XX diagnosis, drug screening, clinical trial monitoring and treatment of
 XX diseases and disorders.

XX Claim 6; Page 54-55; 60pp; English.

XX The invention relates to isolated nucleic acids encoding novel human
 XX transferase proteins (NHP). The nucleic acids and proteins are useful for
 XX diagnosis, drug screening, clinical trial monitoring and treatment of
 XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 XX useful for augmenting the efficacy of chemotherapeutic agents used in
 XX treatment of breast or prostate cancer. The nucleic acid is also useful
 XX in NHP gene regulation, and as antisense primers in amplification
 XX reactions of NHP gene sequences. NHPs are useful for producing
 XX antibodies. The present sequence represents a novel human transferase
 XX which has sequence similarity to N-acetyl-galactosaminyltransferase

XX Sequence 631 AA:

Query Match 98.9%; Score 3242.5; DB 4; Length 631;
 Best Local Similarity 95.4%; Pred. No. 0;
 Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 1 MRKREKRLQVALVLAALVLLPVGIMALYRROPDGTGGGGAAPAGGSGSHSRK 60
 DB 1 MRKREKRLQVALVLAALVLLPVGIMALYRROPDGTGGGGAAPAGGSGSHSRK 60

QY 61 KTFPLDGGOKLQMDHDKKATRRDAORVNGEGGRPYPMTDAERVDAAYRENGFNIVSDK 120
 DB 61 KTFPLDGGOKLQMDHDKKATRRDAORVNGEGGRPYPMTDAERVDAAYRENGFNIVSDK 120

QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNCGMSLRTVHSLNRSPELVAEI 180
 DB 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNCGMSLRTVHSLNRSPELVAEI 180

QY 181 VLVDPSDRHLLKKPLEDYVALTPSVRIILTKKREGILIRRMGASVATDVTTFDSDHC 240
 DB 181 VLVDPSDRHLLKKPLEDYVALTPSVRIILTKKREGILIRRMGASVATDVTTFDSDHC 240

QY 241 EAVVNNLPLLDIARRKTIIVCPMIDVIDHDPFRYETOGDMRGAFDEMYKKIPIIP 300
 DB 241 EAVVNNLPLLDIARRKTIIVCPMIDVIDHDPFRYETOGDMRGAFDEMYKKIPIIP 300

QY 301 PELQKADPSDPFESPVWAGGLFAVDRKMFWEELGYPGGLIENGEOYEISFK----- 352
 DB 301 PELQKADPSDPFESPVWAGGLFAVDRKMFWEELGYPGGLIENGEOYEISFK----- 352

QY 353 -----VMMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARNLK 391
 DB 353 -----VMMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARNLK 391

QY 361 VSNWSDQAVFLPAPNMLALQVMMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARNLK 420
 DB 361 VSNWSDQAVFLPAPNMLALQVMMCGRMEDIPCSRVGHIYRKYVPYKYPAGVSLARNLK 420

QY 392 RVAEVMDEYAEIYORREYHLSAGDAVQKLRSSLNCKSPKFMFKIAMDLPKFPY 451
 DB 421 RVAEVMDEYAEIYORREYHLSAGDAVQKLRSSLNCKSPKFMFKIAMDLPKFPY 480

QY 452 PVEPPAANGETRNVGTGLCADTKGALGSPRLBSCVTRGBAANNNOVFTTREDI 511
 DB 481 PVEPPAANGETRNVGTGLCADTKGALGSPRLBSCVTRGBAANNNOVFTTREDI 540

QY 512 RGGDPQHTKKFCFPAISHSPVTLVYDCHSMKNQMLKRYKDKTLVHPVSGSCMDGSESDH 571
 DB 541 RGGDPQHTKKFCFPAISHSPVTLVYDCHSMKNQMLKRYKDKTLVHPVSGSCMDGSESDH 600

QY 572 RIFMNTCNPSLTLQWLFEHTNSTVLEKEN 601
 DB 601 RIFMNTCNPSLTLQWLFEHTNSTVLEKEN 630

RESULT 5
 AAU07772
 ID AAU07772 standard; protein; 506 AA.
 XX
 AC AAU07772;
 DT 04-DEC-2001 (first entry)
 XX
 DE Human novel transferase protein, NHP #15.
 XX
 KM Human; transferase; breast cancer; prostate cancer; immunogen;
 KM gene therapy; antisense.
 OS Homo sapiens.
 XX
 PN WO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185920P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 XX
 XX (LEXI-) LEXICON GENETICS INC.
 XX
 XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A;
 XX Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
 XX Potter DG;
 XX MPI; 2001-550185/61.
 XX N-PSDB; AAS12624.

XX Novel nucleic acid sequences encoding novel human proteins useful for
 XX diagnosis, drug screening, clinical trial monitoring and treatment of
 XX diseases and disorders.

XX Claim 8; Page 47-48; 60pp; English.

XX The invention relates to isolated nucleic acids encoding novel human
 XX transferase proteins (NHP). The nucleic acids and proteins are useful for
 XX diagnosis, drug screening, clinical trial monitoring and treatment of
 XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 XX useful for augmenting the efficacy of chemotherapeutic agents used in
 XX treatment of breast or prostate cancer. The nucleic acid is also useful
 XX in NHP gene regulation, and as antisense primers in amplification
 XX reactions of NHP gene sequences. NHPs are useful for producing
 XX antibodies. The present sequence represents a novel human transferase
 XX which has sequence similarity to N-acetyl-galactosaminyltransferase

XX Sequence 506 AA:

Query Match 84.5%; Score 2771; DB 4; Length 506;
 Best Local Similarity 100.0%; Pred. No. 8.2e-274;

Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDOAYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYETLPTNTSIIPENEG 157
 DB 1 MTDARVDOAYRENGFNIVYSDKISLNRSLPDIRHPNCSKRYETLPTNTSIIPENEG 60

QY 158 WSSLRLRVSHVSLNRPPELVVAEIVLVDFSDREHLKKPLBEDYMLPPSVRLIRTKKEGL 217
 DB 61 WSSLRLRVSHVSLNRPPELVVAEIVLVDFSDREHLKKPLBEDYMLPPSVRLIRTKKEGL 120

QY 218 IRTFRLGASVATGCVITFLDSHCANVNLPLLDRIARNKTIYVCMIDVIDHDDPRYE 277
 DB 121 IRTFRLGASVATGCVITFLDSHCANVNLPLLDRIARNKTIYVCMIDVIDHDDPRYE 180

QY 278 TQGDARGA FDMEMTKRIPPELOKADSDPFESPVMAGLFAVDRKMFELGSDP 337
 DB 181 TQGDARGA FDMEMTKRIPPELOKADSDPFESPVMAGLFAVDRKMFELGSDP 240

QY 338 GLBIVGGEQYIEISFKVWMCGRMEDIPCSRVGHIYRKVVPYKVPAGVSLARLKRVAEW 397
 DB 241 GLBIVGGEQYIEISFKVWMCGRMEDIPCSRVGHIYRKVVPYKVPAGVSLARLKRVAEW 300

QY 398 MDEYAEIYORRPEYRHLASGDAVAVOKKLRSLSNCKSFKPMFTKIAMDLPRKYPVPEPA 457
 DB 301 MDEYAEIYORRPEYRHLASGDAVAVOKKLRSLSNCKSFKPMFTKIAMDLPRKYPVPEPA 360

QY 458 AAMEIINNVGTCADPRKHALGSPRLLEGVRRGGAANNMVOFTFRREDIRREDPQ 517
 DB 361 AAMEIINNVGTCADPRKHALGSPRLLEGVRRGGAANNMVOFTFRREDIRREDPQ 420

QY 518 HTKFCFDSISHTSPVLYDCHSMKGNQMLWKYRKDTLYHPVSSGCMDCSESDRIEMNT 577
 DB 421 HTKFCFDSISHTSPVLYDCHSMKGNQMLWKYRKDTLYHPVSSGCMDCSESDRIEMNT 480

QY 578 CNPSSLTQOMLFHTNSTVLEKFNRN 603
 DB 481 CNPSSLTQOMLFHTNSTVLEKFNRN 506

RESULT 6
 AAG79781
 ID AAG79781 standard; protein: 506 AA.
 XX
 AC AAG79781;
 XX
 DT 16-APR-2003 (first entry)
 XX
 DE Carbohydrate-associated protein (CHOP) -3.
 XX
 KM Human; carbohydrate associated polypeptides; CHOP; cancer;
 KM carbohydrate metabolism; diabetes; anaemia; hypoglycaemia; obesity;
 KM glycoenzyme storage disease; neurological dysfunction; gene therapy;
 KM cell proliferation; actinic keratosis; arteriosclerosis; inflammation;
 KM atherosclerosis; cirrhosis; hepatitis; psoriasis; autoimmune disorder;
 KM acquired immunodeficiency syndrome; AIDS; humanized; transgenic;
 KM Addison's disease; allergy; asthma; contact dermatitis; bronchitis; gout;
 KM Hashimoto's thyroiditis; infection; reproduction; endometriosis;
 KM polycystic ovary syndrome; ovarian hyperstimulation syndrome;
 KM Down syndrome; cystic fibrosis; sickle cell anaemia; thalassemia;
 KM myocardiitis; cardiomyopathy; Alzheimer's disease; Parkinson's disease;
 KM dementia; depression; epilepsy; Tourette's disorder; schizophrenia;
 KM central nervous system; cerebral palsy; mood; anxiety; knockin.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT 3. .503
 FT /label= Acetylglucosaminyl transferase
 FT /note= "Identified by BLAST_DOMO"
 FT Peptide
 FT 8. .493
 FT /label= Acetylglucosaminyl transferase
 FT /note= "Identified by BLAST_DOMO"
 FT Peptide
 FT 11. .492

FT /label= Acetylglucosaminyl transferase
 FT /note= "Identified by BLAST_DOMO"
 FT Peptide
 FT 51. .236
 FT /label= Glycosyl transferase sequence
 FT /note= "Identified by HMMER_PPAM"
 FT Peptide
 FT 202. .362
 FT /label= N-acetylglucosaminyl transferase
 FT /note= "Identified by BLAST_PRODUM"
 FT Region
 FT 364. .402
 FT /label= QXW lectin repeat
 FT /note= "Identified by HMMER_PPAM"
 FT Region
 FT 417. .455
 FT /label= QXW lectin repeat
 FT /note= "Identified by HMMER_PPAM"
 FT Region
 FT 456. .495
 FT /label= QXW lectin repeat
 FT /note= "Identified by HMMER_PPAM"
 XX
 PN W0200297060-A2.
 XX
 PD 05-DEC-2002.
 XX
 PF 22-MAY-2002; 2002WO-US018354.
 XX
 PR 25-MAY-2001; 2001US-0293768P.
 PR 01-AUG-2001; 2001US-0309548P.
 PR 23-AUG-2001; 2001US-0314400P.
 PR 19-OCT-2001; 2001US-0343706P.
 PR 07-DEC-2001; 2001US-0337999P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Sarnakar A, Goryad AE, Hafalia AJA, Duggan BM, Emerling BM;
 PI Tsou CH, Nguyen DB, Lee EA, Yue H, Forsythe TJ, Li JX;
 PI Thangaveilu K, Walla NK, Burford N, Mason PM, Lal PG, Lee S;
 PI Becha SD, Tang YT;
 XX
 DR WPI: 2003-140462/13.
 DR N-PSDB; ABA00833.
 XX
 PT Novel human carbohydrate associated polypeptide, useful in diagnosis,
 PT treatment and prevention of carbohydrate metabolism, cell proliferative,
 PT autoimmune/inflammatory, reproductive, and neurological disorders.
 XX
 PS Claim 1; Page 126-27; 141pp; English.
 XX
 CC The sequences given in AAG79779-88 represent human carbohydrate
 CC associated polypeptides (CHOP). The CHOP polypeptides and the nucleotide
 CC sequences encoding them, are useful for diagnosing, treating and
 CC preventing cancer, carbohydrate metabolic disorders (e.g. diabetes,
 CC anaemia, hypoglycaemia, obesity, glycogen storage disease, neurological
 CC dysfunction), cell proliferative disorders (e.g. actinic keratosis,
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis),
 CC autoimmune/inflammatory disorders (such as acquired immunodeficiency
 CC syndrome (AIDS), Addison's disease, allergies, asthma, contact
 CC dermatitis, bronchitis, gout, Hashimoto's thyroiditis), viral, bacterial,
 CC fungal, parasitic, protozoal and helminthic infections, reproductive
 CC disorders (e.g. endometriosis, polycystic ovary syndrome, ovarian
 CC hyperstimulation syndrome), genetic disorders (e.g. Down syndrome, cystic
 CC fibrosis, sickle cell anaemia, thalassemia), cardiac disorders (e.g.
 CC myocardiitis, cardiomyopathy), neurological disorders (e.g. Alzheimer's
 CC disease, Parkinson's disease, dementia, depression, epilepsy, Tourette's
 CC disorder, schizophrenia), and developmental disorders of central nervous
 CC system (e.g. cerebral palsy), mental disorders (e.g. mood, anxiety). CHOP
 CC proteins are useful in a number of drug screening techniques, and to
 CC analyse the proteome of a tissue or cell type. CHOP cDNA is useful for
 CC creating "knockin" humanized animals or transgenic animals to model human
 CC diseases, in somatic or germline gene therapy, to generate a transcript
 CC image of a tissue or cell type, for detecting differences in the
 CC chromosomal location due to translocation, inversion, etc., among normal,
 CC carrier or affected individuals, and as hybridization probes for mapping
 CC naturally occurring genomic sequences

SQ Sequence 506 AA;
 Query Match 84.5%; Score 2771; DB 6; Length 506;
 Best Local Similarity 100.0%; Pred. No. 8.2e-274;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 98 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIRHPNCNKKRYLETLPNTSIIIPPHNEG 157
 1 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIRHPNCNKKRYLETLPNTSIIIPPHNEG 60
 DB 158 MSSLLRTVSHVNLRSPELVAEIVLVDFSDREHLKKPLEDYMALPPSVAILRTKKREG 217
 61 MSSLLRTVSHVNLRSPELVAEIVLVDFSDREHLKKPLEDYMALPPSVAILRTKKREG 120
 QY 218 IRTTMLGASVATGVDVITFLDSHCENAVNMLPPLDRIARNRKTIYCPMIDVIDHDDFRYE 277
 121 IRTTMLGASVATGVDVITFLDSHCENAVNMLPPLDRIARNRKTIYCPMIDVIDHDDFRYE 180
 DB 278 TQAGDAMRGAFDWMYYKRIPIPELOKADPSDFESPVMAGLFVADRKKFWELGSDYD 337
 181 TQAGDAMRGAFDWMYYKRIPIPELOKADPSDFESPVMAGLFVADRKKFWELGSDYD 240
 QY 338 GLEIMGEGEYEIFSKVMCGRMEDIPCSRGHYRKYVPYKVPAGVSLARNLKRYAEVW 397
 241 GLEIMGEGEYEIFSKVMCGRMEDIPCSRGHYRKYVPYKVPAGVSLARNLKRYAEVW 300
 DB 398 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCKSFKKFMTKIANDLPKYPVPVPPA 457
 301 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCKSFKKFMTKIANDLPKYPVPVPPA 360
 QY 458 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGGAANNMNVFTTWRREDIRPGDPQ 517
 361 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGGAANNMNVFTTWRREDIRPGDPQ 420
 DB 518 HTKKFCFDASHSPVLYDCHSHKGNQMLKKYRKDKTYHPVSGSCMDCSBDRILPMNT 577
 421 HTKKFCFDASHSPVLYDCHSHKGNQMLKKYRKDKTYHPVSGSCMDCSBDRILPMNT 480
 QY 578 CNPSSLTQOWLFEHTNSTVLEKFNRN 603
 481 CNPSSLTQOWLFEHTNSTVLEKFNRN 506
 DB
 RESULT 7
 AAU07771
 ID AAU07771 standard; protein; 535 AA.
 XX
 AC AAU07771;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Human novel transferase protein, NHP #14.
 XX
 KM Human; transferase; breast cancer; prostate cancer; immunogen;
 KM gene therapy; antilease.
 XX
 OS Homo sapiens.
 XX
 PN MO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001MO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185920P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Hlibun E, Turner CA, Friedrich G, Abulin A;
 PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;
 PI Potter DG;

XX
 DR WPI; 2001-550165/61.
 DR N-PSDB; AAS12618.
 XX
 PT Novel nucleic acid sequences encoding novel human proteins useful for
 PT diagnosis, drug screening, clinical trial monitoring and treatment of
 PT diseases and disorders.
 XX
 PS Claim 7; Page 45-46; 60pp; English.
 XX
 CC The invention relates to isolated nucleic acids encoding novel human
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for
 CC diagnosis, drug screening, clinical trial monitoring and treatment of
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 CC useful for augmenting the efficacy of chemotherapeutic agents used in
 CC treatment of breast or prostate cancer. The nucleic acid is also useful
 CC in NHP gene regulation, and as antisense primers in amplification
 CC reactions of NHP gene sequences. NHPs are useful for producing
 CC antibodies. The present sequence represents a novel human transferase
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase
 CC
 SQ Sequence 535 AA;
 Query Match 83.8%; Score 2746.5; DB 4; Length 535;
 Best Local Similarity 94.6%; Pred. No. 2.9e-271;
 Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
 QY 98 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIRHPNCNKKRYLETLPNTSIIIPPHNEG 157
 1 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIRHPNCNKKRYLETLPNTSIIIPPHNEG 60
 DB 158 MSSLLRTVSHVNLRSPELVAEIVLVDFSDREHLKKPLEDYMALPPSVAILRTKKREG 217
 61 MSSLLRTVSHVNLRSPELVAEIVLVDFSDREHLKKPLEDYMALPPSVAILRTKKREG 120
 QY 218 IRTTMLGASVATGVDVITFLDSHCENAVNMLPPLDRIARNRKTIYCPMIDVIDHDDFRYE 277
 121 IRTTMLGASVATGVDVITFLDSHCENAVNMLPPLDRIARNRKTIYCPMIDVIDHDDFRYE 180
 DB 278 TQAGDAMRGAFDWMYYKRIPIPELOKADPSDFESPVMAGLFVADRKKFWELGSDYD 337
 181 TQAGDAMRGAFDWMYYKRIPIPELOKADPSDFESPVMAGLFVADRKKFWELGSDYD 240
 QY 338 GLEIMGEGEYEIFSKVMCGRMEDIPCSRGHYRKYVPYKVPAGVSLARNLKRYAEVW 397
 241 GLEIMGEGEYEIFSKVMCGRMEDIPCSRGHYRKYVPYKVPAGVSLARNLKRYAEVW 300
 DB 398 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCKSFKKFMTKIANDLPKYPVPVPPA 457
 301 MDEYAEITYORREYRHLASGDAVAVQKKLSLNCKSFKKFMTKIANDLPKYPVPVPPA 360
 QY 458 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGGAANNMNVFTTWRREDIRPGDPQ 517
 361 AANGEIRNVGTGLCADTKHGALGSPRLBECVGRGGAANNMNVFTTWRREDIRPGDPQ 420
 DB 518 HTKKFCFDASHSPVLYDCHSHKGNQMLKKYRKDKTYHPVSGSCMDCSBDRILPMNT 577
 421 HTKKFCFDASHSPVLYDCHSHKGNQMLKKYRKDKTYHPVSGSCMDCSBDRILPMNT 480
 QY 578 CNPSSLTQOWLFEHTNSTVLEKFNRN 603
 481 CNPSSLTQOWLFEHTNSTVLEKFNRN 535
 DB
 RESULT 8
 AAB94733
 ID AAB94733 standard; protein; 339 AA.
 XX
 AC AAB94733;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:15766.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-00116126.
 PF
 XX 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 PA (HELI-) HELIX RES INST.
 XX
 PI Oka T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 PS
 XX Claim 8; SEQ ID NO 15766; 2537bp + Sequence listing; English.
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AA892446 to AA895893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 CC
 SQ Sequence 339 AA;
 QY
 Query Match 57.7%; Score 1893; DB 4; Length 339;
 Best Local Similarity 99.7%; Pred. No. 3.1e-184;
 Matches 338; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MIDVIDDDPFRYEQADAMGAFDMWYRIRIPPELOKADSDPESPVMAGGJPAV 60
 QY 325 DRKFMELGCGYDPLGELTWGEGQYEISFKVMCGGRMEDIPCSRGHYTKRYKVPYKDVAGV 384
 Db 61 DRKFMELGCGYDPLGELTWGEGQYEISFKVMCGGRMEDIPCSRGHYTKRYKVPYKDVAGV 120
 QY 385 SLAANLKRVAEVMDEVAEYIYQRRPEYRHLASGDVAVQKTLRSSLNCKSFKMFMTKIAW 444
 Db 121 SLAANLKRVAEVMDEVAEYIYQRRPEYRHLASGDVAVQKTLRSSLNCKSFKMFMTKIAW 180
 QY 445 DLRFYFVPEPAAAKGEIRNVGTGLCADTKHGLSGPLRLGCGVRGGEAAMNMNQVFT 504

Db 181 DLRFYFVPEPAAAKGEIRNVGTGLCADTKHGLSGPLRLGCGVRGGEAAMNMNQVFT 240
 QY 505 FTWRREDIRPGDPQHTKKFCFDALSHTSPTLLDCHSMKGNOLMKYRKDKTLVHPVSGSCM 564
 Db 241 FTWRREDIRPGDPQHTKKFCFDALSHTSPTLLDCHSMKGNOLMKYRKDKTLVHPVSGSCM 300
 QY 565 DCSSEDRIRFMNTCNPSLTLQWLPEHTNSTVLEKFN 603
 Db 301 DCSSEDRIRFMNTCNPSLTLQWLPEHTNSTVLEKFN 339
 RESULT 9
 AAU07776
 ID AAU07776 standard; protein; 366 AA.
 XX
 AC AAU07776;
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE Human novel transferase protein, NHP #19.
 XX
 KW Human; transferase; breast cancer; prostate cancer; immunogen;
 KW gene therapy; antisense.
 XX
 OS Homo sapiens.
 XX
 PN WO200164903-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US006460.
 XX
 PR 29-FEB-2000; 2000US-0185320P.
 PR 02-MAR-2000; 2000US-0186558P.
 PR 24-MAR-2000; 2000US-0191849P.
 XX
 PA (LEXI-) LEXICON GENETICS INC.
 XX
 PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;
 PI Zambrowicz B, Sands AT, Walke DW, Wilgowski NL, Hu Y, Kieke JA;
 PI Potter DG;
 DR WPI; 2001-550185/61.
 DR N-PSDB; AAS12623.
 XX
 PT Novel nucleic acid sequences encoding novel human proteins useful for
 PT diagnosis, drug screening, clinical trial monitoring and treatment of
 PT diseases and disorders.
 XX
 PS Disclosure; Page 52-53; 60pp; English.
 XX
 CC The invention relates to isolated nucleic acids encoding novel human
 CC transferase proteins (NHP). The nucleic acids and proteins are useful for
 CC diagnosis, drug screening, clinical trial monitoring and treatment of
 CC diseases and disorders e.g. breast and prostate cancer. NHPs can also be
 CC useful for augmenting the efficacy of chemotherapeutic agents used in
 CC treatment of breast or prostate cancer. The nucleic acid is also useful
 CC in NHP gene regulation, and as antisense primers in amplification
 CC reactions of NHP gene sequences. NHPs are useful for producing
 CC antibodies. The present sequence represents a novel human transferase
 CC which has sequence similarity to N-acetyl-galactosaminyltransferase
 CC
 SQ Sequence 366 AA;
 QY
 Query Match 57.3%; Score 1877; DB 4; Length 366;
 Best Local Similarity 100.0%; Pred. No. 1.5e-182;
 Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 1 MRRREKRLQAVALLVLAALVLLPNVGMALYRERQDPGTGSGGAAYAPAGGSGSHRQK 60
 1 MRRREKRLQAVALLVLAALVLLPNVGMALYRERQDPGTGSGGAAYAPAGGSGSHRQK 60

QY 61 KTEFLDGGQKLDKMDHKEAIRDAQGVNGEGQRPYPMTDABRVDAQYRENGFNIVYSDK 120
 DB 61 KTEFLDGGQKLDKMDHKEAIRDAQGVNGEGQRPYPMTDABRVDAQYRENGFNIVYSDK 120
 QY 121 ISLNRLPDIIRHNCNCKRYLETLPNTSIIIPHNEMSSLLRTYHSVLRNSPPELVAEI 180
 DB 121 ISLNRLPDIIRHNCNCKRYLETLPNTSIIIPHNEMSSLLRTYHSVLRNSPPELVAEI 180
 QY 181 VLVDSDSDREHLKKPLEDYVALFPYSVILRTKKREGIIRTRMLGASVATGDTVTFDLSHC 240
 DB 181 VLVDSDSDREHLKKPLEDYVALFPYSVILRTKKREGIIRTRMLGASVATGDTVTFDLSHC 240
 QY 241 EANNVNLPLLDRIARRKTIIVCPMIDVIDHDDFRYETQAGAMRGAFDMEYKRIPIR 300
 DB 241 EANNVNLPLLDRIARRKTIIVCPMIDVIDHDDFRYETQAGAMRGAFDMEYKRIPIR 300
 QY 301 PELQKADPSDPSPVWAGGLFVADRKKFWEIGYDGLGIEINGEOYEISFKV 353
 DB 301 PELQKADPSDPSPVWAGGLFVADRKKFWEIGYDGLGIEINGEOYEISFKV 353

RESULT 10

AA62600
 ID AA62600 standard; protein; 407 AA.

AC AA62600;
 XX
 DT 06-SEP-2001 (first entry)

DE Human acetyl galactosyl transferase 45.

XX Human; acetyl galactosyl transferase 45; GalNAc-T45; cancer; haemopathy;
 KM HIV infection; immunological disease; inflammation; gene therapy.

XX Homo sapiens.

PN WO200138545-A1.

PD 31-MAY-2001.

PF 20-NOV-2000; 2000MO-CN000473.

PR 24-NOV-1999; 99CN-00124100.

PA (BIO-R-) BIOROAD GENE DEV LTD SHANGHAI.

PI Mao Y, Xie Y;

DR WPI; 2001-355942/37.

DR N-PSDB; AAA45701.

XX Human acetyl galactosyl transferase 45 and encoded polynucleotide, used
 PT in diagnosis and treatment of malignant tumors, hemopathy, human
 PT immunodeficiency virus infection, immunological diseases and
 PT inflammation.

PS Claim 1; Page 20-21; 33pp; Chinese.

XX The present invention provides the protein and coding sequences of the
 CC human acetyl galactosyl transferase 45 (GalNAc-T45). The sequences can be
 CC used in the treatment of cancer, haemopathy, HIV infection, immunological
 CC diseases and inflammation. The present sequence is the protein of the
 CC invention

XX Sequence 407 AA.

Query Match 55.5%; Score 1818; DB 4; Length 407;

Best Local Similarity 78.5%; Pred. No. 2e-176;

Matches 317; Conservative 36; Mismatches 51; Indels 0; Gaps 0;

QY 200 MALPSPRIIRTKKREGLIRTRMLGASVANGDVITFDSCGANVNLPLLDRIARRNK 259
 DB 1 MARFSKRIIVRTKREGLIRTRMLGASVANGDVITFDSCGANVNLPLLDRIARRNK 60

QY 260 TIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEYKRIPIPELQKADPSDPSPVWAG 319
 DB 61 TIVCPMIDVIDHDDFRYETQAGDAMRGAFDMEYKRIPIPELQKADPSDPSPVWAG 120
 QY 320 GLFVADRKKFWEIGYDGLGIEINGEOYEISFVWVWCGGEMEDVPCSRVGHITRYKVPYK 379
 DB 121 GLFVADRKKFWEIGYDGLGIEINGEOYEISFVWVWCGGEMEDVPCSRVGHITRYKVPYK 180
 QY 380 VPAGVSLARNLKRVAVAEVWDEVAEYIYORRPEYRHLASDVAVQKLRSLNCKSPKWF 439
 DB 181 VPSGTSILARNLKRVAVAEVWDEVAEYIYORRPEYRHLASDVAVQKLRSLNCKSPKWF 240
 QY 440 TKIAMDLPKRYPPVEPPAAWGEIRNVGTGLCADTGHGALGSPRLGEGVGRGEAAMN 499
 DB 241 AAVAMDVPKYYPPVEPPAAWGEIRNVANLGVDSHGATGTFLRDI CVKDGSERTW 300
 QY 500 MQVFTTWMEDIRPGDPQTKKCFPAIHTSPTLYDCHSKGNQJMKYRDKTLYHVP 559
 DB 301 EQLFTFGWREDIRPGDPQTKKCFPAIHTSPTLYDCHSKGNQJMKYRDKTLYHVP 360
 QY 560 SGSCMDCSDRHIFMNTGPNPSLTDQWLFETHNSTVLEKFN 603
 DB 361 SNSCMDCNPAKKIFMARCDPLSETQWLFETHNSTVLEKFN 404

RESULT 11

ABG08187
 ID ABG08187 standard; protein; 319 AA.

AC ABG08187;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8187.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS72374.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostic, forensic, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 20; SEQ ID NO 38546; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 319 AA;

Query Match 42.8%; Score 1403.5; DB 4; Length 319;
 Best Local Similarity 90.8%; Pred. No. 3.5e-134;
 Matches 258; Conservative 4; Mismatches 19; Indels 3; Gaps 2;

QY 322 FAVDRKFWELGSDPGLLEIWCSEGEYSFVWVGGRMEDIPCSRGHYRKVPYKVP 381
 DB 12 FAEAKKCCMP-GGMDLEIAETSSSLFLAMQVMWCGRMEDIPCSRGHYRKVPYKVP 70
 QY 382 AGVSLAR--NLKRYAEVWMDYAEYIYORREYHLSAGDVAVOKKLRSSLNCKSFPMFM 439
 DB 71 AGVSLARSKLKRVAEVMDEYAEYIYORREYHLSAGDVAVOKKLRSSLNCKSFPMFM 130
 QY 440 TKIAMDLPKFYPPVPPPAAMGEIRNVGTGICADTKHGALGSPRLRGCGRGEAAMNN 499
 DB 131 TKIAMDLPKFYPPVPPPAAMGEIRNVGTGICADTKHGALGSPRLRGCGRGEAAMNN 190
 QY 500 MOVFTFWREDIRGDDQHTKKCFDAISHTSPVTLVDCSHMKNOQMKRKDKTLVHPV 559
 DB 191 MOVFTFWREDIRGDDQHTKKCFDAISHTSPVTLVDCSHMKNOQMKRKDKTLVHPV 250
 QY 560 SGSCMDCSESDHRIFNNTCNPSLSLTQOVLFEHTNSTVLEKNNRN 603
 DB 251 SGSCMDCSESDHRIFNNTCNPSLSLTQOVLFEHTNSTVLEKNNRN 294

RESULT 12
 ADE25865
 ID ADE25865 standard; protein; 275 AA.

AC ADE25865;

XX 29-JAN-2004 (first entry)

DE Human N-acetylglucosaminyl transferase T10.

XX CYTOSTATIC; respiratory; antiasthmatic; antiinflammatory; endocrine;
 XX immunosuppressive; antidiarrhetic; antibacterial; vulnery;
 XX N-acetylglucosaminyl transferase; GalNAc-4-epimerase; GalNAc-4-epimerase;
 XX GalNAc-4-epimerase; GalNAc-4-epimerase; GalNAc-4-epimerase;
 XX lung disease; mucous accumulation; asthma; chronic bronchitis;
 XX smoker's lung; cystic fibrosis; exocrine gland disease;
 XX mucin secretion decrease; Sjogren's syndrome; dry mouth;
 XX selection-mediated leukocyte trafficking; autoimmunity; arthritis;
 XX leukaemia; lymphoma; immunosuppression; sepsis; wound healing;
 XX inflammation; mucin secretion inhibition; hypersecretion;
 XX mucin accumulation; chronic obstructive respiratory pulmonary disease;
 XX asthma; lectin domain inhibition; GalNAc-4-epimerase;
 XX GalNAc-4-epimerase; GalNAc-4-epimerase; GalNAc-4-epimerase;
 XX core 1 beta1,3-galactosyltransferase; alpha2,6-sialyltransferase;
 XX N-acetylglucosaminyl transferase 4AP24 glycopeptide; O-glycosylation;
 XX breast cancer; human; GalNAcT10.

OS Homo sapiens.

XX US2003186850-A1.

XX 02-OCT-2003.

XX 12-NOV-2002; 2002US-00292896.

XX 11-MAY-2000; 2000US-0203331P.

XX 10-MAY-2001; 2001WO-DK000328.

XX 08-NOV-2001; 2002US-0425204P.

XX (GLYC-) GLYCOZYME APS.

XX Clausen H, Bennett EP, Haasean H, Reis CA;

XX WPI; 2003-831259/77.

XX Modulating N-acetylglucosaminyltransferase functions comprises

XX administration of an N-acetylglucosaminyltransferase lectin domain

XX inhibitor.

XX Disclosure; Page 11; 65pp; English.

XX The invention describes a method of modulating polypeptide N-
 CC acetylglucosaminyl (GalNAc)-transferase functions comprising
 CC administration of a GalNAc-transferase lectin domain inhibitor (I). (I)
 CC is used for modulating functions of GalNAc-transferase (e.g. GalNAc-T4,
 CC GalNAc-T7, GalNAc-T2, and GalNAc-T3); for inhibiting at least one lectin
 CC domain of the GalNAc-transferase and modulating the function mediated by
 CC the domain for preparing a medicament for treating tumours and cancers,
 CC lung diseases associated with mucous accumulation (e.g. asthma, chronic
 CC bronchitis, smoker's lung, and cystic fibrosis), diseases of exocrine
 CC glands associated with increased or decreased mucin secretion (e.g.
 CC Sjogren's syndrome and dry mouth), disorders associated with
 CC dysregulation of selectin-mediated leukocyte trafficking (e.g.
 CC autoimmunity, arthritis, leukaemia, lymphoma, immunosuppression,
 CC sepsis, wound healing, acute and chronic inflammation); inhibiting mucin
 CC secretion; and inhibiting hypersecretion and accumulation of mucin in the
 CC lungs of mammals (e.g. humans) suffering from chronic obstructive
 CC respiratory pulmonary diseases, asthma, and cystic fibrosis. (I) is
 CC effective in inhibiting at least one lectin domain of the GalNAc-
 CC transferase and modulating the inhibited function mediated by the domain
 CC (preferably GalNAc-glycopeptide, GalNAc-4-epimerase, or UDP-Gal donor substrate
 CC specificity). (I) selectively inhibits at least one member of the GalNAc-
 CC transferase family without inhibiting other glycosyltransferases
 CC (preferably core 1 beta1,3-galactosyltransferase, alpha2,6-
 CC sialyltransferase, or glycosyltransferase functioning in the O-
 CC glycosylation pathway). This is the amino acid sequence of human N-
 CC acetylglucosaminyl transferase T10.

XX Sequence 275 AA;

Query Match 42.6%; Score 1396; DB 7; Length 275;
 Best Local Similarity 93.7%; Pred. No. 1.6e-133;
 Matches 253; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 336 DPGLEIWCSEGEYSFVWVGGRMEDIPCSRGHYRKVPYKVPYKVPYKVP 393
 DB 6 DGELEIWCSEGEYSFVWVGGRMEDIPCSRGHYRKVPYKVPYKVPYKVPYKVP 65
 QY 394 AEVWMDYAEYIYORREYHLSAGDVAVOKKLRSSLNCKSFPMFMFKIAMDLPKFYPPV 453
 DB 66 AEVWMDYAEYIYORREYHLSAGDVAVOKKLRSSLNCKSFPMFMFKIAMDLPKFYPPV 125
 QY 454 EPPPAAMGEIRNVGTGICADTKHGALGSPRLRGCGRGEAAMNNQVFTFWREDIRP 513
 DB 126 EPPPAAMGEIRNVGTGICADTKHGALGSPRLRGCGRGEAAMNNQVFTFWREDIRP 185
 QY 514 GDPQHTKKCFDAISHTSPVTLVDCSHMKNOQMKRKDKTLVHPVSGSCMDCSESDHRI 573
 DB 186 GDPQHTKKCFDAISHTSPVTLVDCSHMKNOQMKRKDKTLVHPVSGSCMDCSESDHRI 245
 QY 574 FNNTCNPSLSLTQOVLFEHTNSTVLEKNNRN 603
 DB 246 FNNTCNPSLSLTQOVLFEHTNSTVLEKNNRN 275

RESULT 13
ID ADO17613 standard; protein: 276 AA.
XX ADO17613;
XX ADO17613;
DT 26-AUG-2004 (first entry)
XX
XX Human soft tissue sarcoma-upregulated protein - SEQ ID 430.
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
XX Homo sapiens.
XX
XX WO2004048938-A2.
XX
XX 10-JUN-2004.
XX
XX 26-NOV-2003; 2003WO-US038193.
XX
XX 26-NOV-2002; 2002US-0429739P.
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
XX
XX Aziz N, Gineburg WM, Zlotnick A;
XX
XX WPI; 2004-441208/41.
XX
XX Example 2; SEQ ID NO 430; 210bp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX CC which comprises obtaining a first soft tissue sample from an individual
XX CC and a normal soft tissue sample from the same or different individual,
XX CC determining the expression of a gene in both samples and comparing the
XX CC expression of the gene in both soft tissue samples, where a higher level
XX CC of protein expression in the first soft tissue sample indicates the
XX CC presence of soft tissue sarcoma. The method of the invention has
XX CC cytostatic applications and may be useful for detecting soft tissue
XX CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX CC acid sequences may be useful in diagnostic and screening applications.
XX CC The current sequence is that of a human soft tissue sarcoma-upregulated
XX CC protein of the invention. The current sequence is not shown within the
XX CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 276 AA;
Query Match 42.5%; Score 1392; DB 8; Length 276;
Best Local Similarity 97.3%; Pred. No. 4.2e-133;
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;
QY 349 ISFVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA--RNLKRYAEVMMDEYAYIY 406
DB 20 LAAQVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA--RNLKRYAEVMMDEYAYIY 79
QY 407 QRRREYHLSAGDVAOVKKLRSSLNCKSPKFMFTKIAMDLPKFYPPVPEPAAWGEIRNV 466
DB 80 QRRREYHLSAGDVAOVKKLRSSLNCKSPKFMFTKIAMDLPKFYPPVPEPAAWGEIRNV 139
QY 467 GTGICADTKRGALGSPRLREGCVARGREAAAMNNQVTFTRWREIRRGDQHTKKCFDA 526
DB 140 GTGICADTKRGALGSPRLREGCVARGREAAAMNNQVTFTRWREIRRGDQHTKKCFDA 199
QY 527 ISHTSPVTLVDCHSMKGNOLMKYRKDXTLYHPVSGSCMDSESDDHRIFFMTCPNSSLTQQ 586
DB 200 ISHTSPVTLVDCHSMKGNOLMKYRKDXTLYHPVSGSCMDSESDDHRIFFMTCPNSSLTQQ 259
QY 587 WLFHTNSTVLEKFRN 603

DB 260 WLFHTNSTVLEKFRN 276
RESULT 14
ID ABG08189 standard; protein: 335 AA.
XX ABG08189;
XX ABG08189;
DT 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #8180.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS72376.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX
XX Claim 20; SEQ ID NO 38548; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probe, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for creating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 335 AA;
Query Match 42.3%; Score 1385.5; DB 4; Length 335;
Best Local Similarity 87.5%; Pred. No. 2.7e-132;
Matches 253; Conservative 0; Mismatches 1; Indels 35; Gaps 1;
QY 350 SFRVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA--RNLKRYAEVMMDEYAYIYOR 409
DB 47 SFRVMMCGGRMEDIPCSRVGHIYRKVPYKVPAGVSLA--RNLKRYAEVMMDEYAYIYOR 106

QY 410 PEYRHLASGVAVOKKRLSSLNCKSPKEMTKIAMDLPKPYPPVPPAAAMGE----- 462
 DB 107 PEYRHLASGVAVOKKRLSSLNCKSPKEMTKIAMDLPKPYPPVPPAAAMGEQROSM 166
 QY 463 -----INNVGTGLCADTKHGALGSPRLLEGCVGRGE 494
 DB 167 YSKGFMSPRRFGQWVGPAPSPGPGMGINVGTGLCADTKHGALGSPRLLEGCVGRGE 226
 QY 495 AANNNOVFTPTREDIRPDPOHTKKECFDAISHTSPVTLVYDCHSKNGQOLMKYRKDT 554
 DB 227 AANNNOVFTPTREDIRPDPOHTKKECFDAISHTSPVTLVYDCHSKNGQOLMKYRKDT 286
 QY 555 LYHPVSSGSCMDCESDHRIEMNTCNPSLTOQWLFETNSTVLEKEFRN 603
 DB 287 LYHPVSSGSCMDCESDHRIEMNTCNPSLTOQWLFETNSTVLEKEFRN 335

RESULT 15

AAU07770
 ID AAU07770 standard; protein; 269 AA.

AC AAU07770;

DT 04-DEC-2001 (first entry)

DE Human novel transferase protein, NHP #13.

KM Human; transferase; breast cancer; prostate cancer; immunogen;

OS Homo sapiens.

PN W0200164903-A2.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001MO-US006460.

PR 29-FEB-2000; 2000US-0185920P.

PR 02-MAR-2000; 2000US-0186558P.

PR 24-MAR-2000; 2000US-0191849P.

XX (LEXI-) LEXICON GENETICS INC.

PI Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A;

PI Zambrowicz B, Sands AT, Walke DW, Wilganowski NL, Hu Y, Kieke JA;

PI Potter DG;

DR WPI; 2001-550185/61.

DR N-PSDB; AAS12617.

XX Novel nucleic acid sequences encoding novel human proteins useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders.

PS Disclosure; Page 44-45; 60pp; English.

XX The invention relates to isolated nucleic acids encoding novel human

XX transferase proteins (NHP). The nucleic acids and proteins are useful for

XX diagnosis, drug screening, clinical trial monitoring and treatment of

XX diseases and disorders e.g. breast and prostate cancer. NHPs can also be

XX useful for augmenting the efficacy of chemotherapeutic agents used in

XX treatment of breast or prostate cancer. The nucleic acid is also useful

XX in NHP gene regulation, and as antisense primers in amplification

XX CC reactions of NHP gene sequences. NHPs are useful for producing

XX CC antibodies. The present sequence represents a novel human transferase

XX which has sequence similarity to N-acetyl-galactosaminyltransferase

XX SQ Sequence 269 AA;

Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 98 MTDARVDQAYRNGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157
 DB 1 MTDARVDQAYRNGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60
 QY 158 WSLRRTVHSVLRSPPELVAAEIVLVDPFSDREHLKKPLEDYVALPFSYRIITKKREG 217
 DB 61 WSLRRTVHSVLRSPPELVAAEIVLVDPFSDREHLKKPLEDYVALPFSYRIITKKREG 120
 QY 218 IRTRLGASVATGDTITFLDSHCANVNLPLLDRIANRRTIVCPMIDVIDHDPFR 277
 DB 121 IRTRLGASVATGDTITFLDSHCANVNLPLLDRIANRRTIVCPMIDVIDHDPFR 180
 QY 278 TQGDAMRGAFFDEMYTKRIPPELOKADPSDPFSPVMAAGLFAVDRKWFELGSDP 337
 DB 181 TQGDAMRGAFFDEMYTKRIPPELOKADPSDPFSPVMAAGLFAVDRKWFELGSDP 240
 QY 338 GLEIWGGEQYEISFKV 353
 DB 241 GLEIWGGEQYEISFKV 256

Search completed: December 20, 2004, 14:09:34
 Job time: 164 secs

Query March 41.8%; Score 1370; DB 4; Length 269;
 Best Local Similarity 100.0%; Pred. No. 7.2e-131;

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2004, 15:26:33 ; Search time 40 Seconds

(without alignments)
999.744 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278
Sequence: 1 MRKKRRLQALVALVLAIV.....TQGMLEHTNSTYLEKRN 603

Scoring table:

BIOSDM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	4	US-09-795-926-43
2	3242.5	98.9	631	4	US-09-795-926-41
3	2771	84.5	506	4	US-09-795-926-31
4	2746.5	83.8	535	4	US-09-795-926-39
5	1877	57.3	366	4	US-09-795-926-39
6	1370	41.8	269	4	US-09-795-926-27
7	1311	40.0	321	4	US-09-795-926-35
8	1115	34.0	517	4	US-08-967-508-19
9	1115	34.0	517	3	US-08-967-508-19
10	1115	34.0	517	3	PCT-US94-02552-19
11	1115	34.0	539	2	US-08-967-508-9
12	1115	34.0	559	3	US-08-967-506-9
13	1115	34.0	559	5	PCT-US94-02552-9
14	1026	31.3	578	4	US-09-217-306B-2
15	1024	31.2	560	4	US-09-217-306B-10
16	990	30.2	240	4	US-09-795-926-33
17	988	30.1	209	4	US-09-795-926-37
18	971	29.6	552	4	US-10-140-002-196
19	941	28.7	302	4	US-09-270-767-33733
20	926	28.2	639	3	US-09-347-488-2
21	920	28.1	638	3	US-09-347-488-2
22	916	27.9	633	2	US-08-648-298-2
23	804	24.5	224	4	US-09-795-926-23
24	744	22.7	603	4	US-09-973-457-2
25	675	20.6	289	4	US-09-270-767-45334
26	561	17.1	238	4	US-09-270-767-46411
27	503.5	15.4	366	4	US-09-270-767-43543

28	483	14.7	143	4	US-09-795-926-21	Sequence 21, Appl
29	481	14.7	112	4	US-09-795-926-25	Sequence 25, Appl
30	450.5	13.7	149	4	US-09-973-457-6	Sequence 6, Appl
31	428.5	13.1	161	4	US-09-270-767-58910	Sequence 58910, A
32	287	8.8	148	4	US-09-513-996C-7570	Sequence 7570, Ap
33	277.5	8.5	122	4	US-09-270-767-32872	Sequence 32872, A
34	241.5	7.4	313	4	US-09-270-767-46209	Sequence 46209, A
35	234.5	7.2	300	4	US-09-270-767-43809	Sequence 43809, A
36	228.5	7.0	320	4	US-09-270-767-61773	Sequence 61773, A
37	218.5	6.7	72	4	US-09-270-767-61989	Sequence 61989, A
38	168	5.1	73	4	US-09-270-767-33350	Sequence 33350, A
39	167	5.1	73	4	US-09-270-767-33441	Sequence 33441, A
40	167	5.1	101	4	US-09-270-767-48658	Sequence 48658, A
41	167	5.1	101	4	US-09-134-000C-5947	Sequence 5947, Ap
42	133	4.1	424	4	US-09-270-767-33601	Sequence 33601, A
43	129	3.9	88	4	US-09-270-767-48818	Sequence 48818, A
44	129	3.9	88	4	US-09-973-457-5	Sequence 5, Appl
45	128	3.9	135	4	US-09-328-352-4416	Sequence 4416, Ap
46	127	3.9	436	4	US-09-437-277-3	Sequence 3, Appl
47	125.5	3.8	965	4	US-09-437-277-1	Sequence 1, Appl
48	124	3.8	702	4	US-09-134-000C-5087	Sequence 5087, Ap
49	123.5	3.8	534	4	US-09-107-532A-6889	Sequence 6889, Ap
50	123	3.8	721	4	US-09-134-000C-5086	Sequence 5086, Ap
51	121.5	3.7	1056	4	US-08-468-812-4	Sequence 4, Appl
52	119.5	3.6	492	2	US-08-468-812-7	Sequence 7, Appl
53	119.5	3.6	492	3	US-08-590-563-4	Sequence 4, Appl
54	119.5	3.6	492	3	US-08-590-563-7	Sequence 7, Appl
55	119.5	3.6	492	3	US-09-770-621-4	Sequence 4, Appl
56	119.5	3.6	492	4	US-09-770-621-7	Sequence 7, Appl
57	119.5	3.6	492	4	US-09-235-832-7	Sequence 7, Appl
58	119.5	3.6	492	4	US-09-235-832-4	Sequence 4, Appl
59	119.5	3.6	492	2	US-08-468-812-5	Sequence 5, Appl
60	118	3.6	480	2	US-08-590-563-5	Sequence 5, Appl
61	118	3.6	480	3	US-09-770-621-5	Sequence 5, Appl
62	118	3.6	480	4	US-09-235-832-5	Sequence 5, Appl
63	118	3.6	480	4	US-09-270-767-33890	Sequence 33890, A
64	116.5	3.6	88	4	US-09-270-767-49107	Sequence 49107, A
65	116.5	3.6	909	3	US-09-425-853-2	Sequence 2, Appl
66	112.5	3.4	491	2	US-08-468-812-8	Sequence 8, Appl
67	111.5	3.4	491	3	US-08-590-563-8	Sequence 8, Appl
68	111.5	3.4	491	4	US-09-770-621-8	Sequence 8, Appl
69	111.5	3.4	491	4	US-09-235-832-8	Sequence 8, Appl
70	111.5	3.4	491	4	US-09-130-337A-25	Sequence 25, Appl
71	107.5	3.3	270	4	US-09-495-406-25	Sequence 25, Appl
72	107	3.3	270	4	US-09-816-028A-39	Sequence 39, Appl
73	107	3.3	270	4	US-10-303-162-32	Sequence 32, Appl
74	107	3.3	187	4	US-09-543-681A-6535	Sequence 6535, Ap
75	105.5	3.2	336	4	US-09-330-945-39	Sequence 39, Appl
76	105.5	3.2	127	1	US-08-597-236-10	Sequence 10, Appl
77	105	3.2	324	1	US-08-746-882A-10	Sequence 10, Appl
78	105	3.2	324	1	US-08-378-761A-77	Sequence 77, Appl
79	105	3.2	540	1	US-08-485-886-77	Sequence 77, Appl
80	105	3.2	540	1	US-09-583-110-4849	Sequence 4849, Ap
81	102.5	3.1	317	4	US-08-635-552A-4	Sequence 4, Appl
82	102.5	3.1	395	4	US-08-270-767-3341	Sequence 3341, Appl
83	102	3.1	419	2	US-09-146-893-2	Sequence 2, Appl
84	101.5	3.1	419	4	US-08-675-499A-5	Sequence 5, Appl
85	101.5	3.1	419	4	US-08-812-008-5	Sequence 5, Appl
86	101.5	3.1	328	4	US-09-583-110-3554	Sequence 3554, Ap
87	101.5	3.1	331	4	US-09-634-238-247	Sequence 247, Appl
88	101.5	3.1	331	4	US-09-495-406-17	Sequence 17, Appl
89	101	3.1	303	4	US-09-816-028A-29	Sequence 29, Appl
90	100	3.0	303	4	US-10-303-162-22	Sequence 22, Appl
91	99.5	3.0	341	4	US-09-495-406-15	Sequence 15, Appl
92	99.5	3.0	301	4	US-09-816-028A-27	Sequence 27, Appl
93	99	3.0	301	4	US-10-303-162-27	Sequence 27, Appl
94	98	3.0	623	1	US-08-653-740-7	Sequence 7, Appl
95	98	3.0	623	2	US-09-275-925-7	Sequence 7, Appl
96	98	3.0	623	3		
97	98	3.0				
98	96.5	2.9				
99	96.5	2.9				
100	96.5	2.9				

ALIGNMENTS

```
RESULT 1
US-09-795-926-43
; Sequence 43, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sander, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ. ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 43
; LENGTH: 603
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-43

Query Match      100.0%; Score 3278; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MRREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60
DB      1 MRREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60

QY      61 KTFELDGQQLKQMDKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
DB      61 KTFELDGQQLKQMDKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120

QY      121 ISLNRSIPDTRHNCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAEI 180
DB      121 ISLNRSIPDTRHNCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAEI 180

QY      181 VLVDDFSDREHLKKPLEDYVALFPSSVILRTKKREGILITRMIGASVATGDTVTFDLSHC 240
DB      181 VLVDDFSDREHLKKPLEDYVALFPSSVILRTKKREGILITRMIGASVATGDTVTFDLSHC 240

QY      241 EAVVNMVLPILLRIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIP 300
DB      241 EAVVNMVLPILLRIARRKTIIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMYYKRIPIP 300

QY      301 PELQKADSPDPESPFWAAGGLFAVDRKFWELGSDYDGLSIWGGEQYEIFSKYMMCGGRM 360
DB      301 PELQKADSPDPESPFWAAGGLFAVDRKFWELGSDYDGLSIWGGEQYEIFSKYMMCGGRM 360

QY      361 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNLKRVAVNMDEYAEYTYGRREBYRHLISAGDV 420
DB      361 EDIPCSRVGHIYRYKYVPYKYPAGVSLARNLKRVAVNMDEYAEYTYGRREBYRHLISAGDV 420
```

```
QY      421 AVQKLRSSLNCKSPFWMTKIAMDLPKFYPPVEPPAAAMGEIRNVGTGLCADTKGALG 480
DB      421 AVQKLRSSLNCKSPFWMTKIAMDLPKFYPPVEPPAAAMGEIRNVGTGLCADTKGALG 480

QY      481 SPLRLGCVRGREGAAMNNQVFTFWREDIRGDPQHTKKCFDAISHTSPVTLVDCHS 540
DB      481 SPLRLGCVRGREGAAMNNQVFTFWREDIRGDPQHTKKCFDAISHTSPVTLVDCHS 540

QY      541 MKGNQLMKRYKQDITLYHPVSGSCMDCSDHRIFNATCNPSLITQOMLEHTNSTVLEKF 600
DB      541 MKGNQLMKRYKQDITLYHPVSGSCMDCSDHRIFNATCNPSLITQOMLEHTNSTVLEKF 600

QY      601 NRN 603
DB      601 NRN 603

RESULT 2
US-09-795-926-41
; Sequence 41, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sander, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; NUMBER OF SEQ. ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-41

Query Match      98.9%; Score 3242.5; DB 4; Length 631;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY      1 MRREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60
DB      1 MRREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAAGGSHSRQK 60

QY      61 KTFELDGQQLKQMDKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120
DB      61 KTFELDGQQLKQMDKEAIRDAQVNGEQGRPYMTDAERVDQAYRENGFNIVYSDK 120

QY      121 ISLNRSIPDTRHNCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAEI 180
DB      121 ISLNRSIPDTRHNCNSKRYLETLPNTSIIIPPHNEGWSLLRTVSHVLRSPPELVAEI 180

QY      181 VLVDDFSDREHLKKPLEDYVALFPSSVILRTKKREGILITRMIGASVATGDTVTFDLSHC 240
DB      181 VLVDDFSDREHLKKPLEDYVALFPSSVILRTKKREGILITRMIGASVATGDTVTFDLSHC 240
```


QY 241 EANNVNLPLLDRIANRRTIYVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKRIPR 300
DB 241 EANNVNLPLLDRIANRRTIYVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKRIPR 300
QY 301 PELQKADPSDPFSPVWAGGLFVADRKFWELGSDYDGLIWSGEQYEISFK----- 352
DB 301 PELQKADPSDPFSPVWAGGLFVADRKFWELGSDYDGLIWSGEQYEISFKGLHMLPRL 360
QY 353 -----VMMCGRMEDIPCRRGHIYRKVYPPKVPAGVSLARNLK 391
DB 361 VSNISMPQAVFLPPAPNNMALQVMMCGRMEDIPCRRGHIYRKVYPPKVPAGVSLARNLK 420
QY 392 RVAEVMWDEYAEYIYORREYRHLISAGDVAVQKLRSSLNCKSPFKMFMTKIAMDLPKFPY 451
DB 421 RVAEVMWDEYAEYIYORREYRHLISAGDVAVQKLRSSLNCKSPFKMFMTKIAMDLPKFPY 480
QY 452 PVEPPAAWGEIRNVGTGLCADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDI 511
DB 481 PVEPPAAWGEIRNVGTGLCADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDI 540
QY 512 RPDDPQHTKFCFCDATISHTSPVTLVYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDSESDDH 571
DB 541 RPDDPQHTKFCFCDATISHTSPVTLVYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDSESDDH 600
QY 572 RIFMNTCNPSLTOQWLFHTNSTVLEKFN 601
DB 601 RIFMNTCNPSLTOQWLFHTNSTVLEKFN 630

RESULT 3
US-09-795-926-31
Sequence 31, Application US/09795926
Patent No. 6555669
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 506
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-31

Query Match 84.5%; Score 2771; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 2,7e-289; Indels 0; Gaps 0;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDQAYRENGFIYVSDKISLNRSLPDIRHPNCNKRYLETLPTNTSIIPFNEG 157
DB 1 MTDARVDQAYRENGFIYVSDKISLNRSLPDIRHPNCNKRYLETLPTNTSIIPFNEG 60

QY 158 WSSLRTVHSVLNRSBPPELVAEIVLVDDPSDEHLLKKPLEDYVALPSPVRIILTKRREGI 217
DB 61 WSSLRTVHSVLNRSBPPELVAEIVLVDDPSDEHLLKKPLEDYVALPSPVRIILTKRREGI 120
QY 218 ITRMLGASVATGCVITPLDISHCEANVNLPLLDRIANRRTIYVCPMIDVIDHDDFRYE 277
DB 121 ITRMLGASVATGCVITPLDISHCEANVNLPLLDRIANRRTIYVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFPDWEYTKRIPRPELOKADPSDPFSPVWAGGLFVADRKFWELGSDYD 337
DB 181 TQAGDMRGAFPDWEYTKRIPRPELOKADPSDPFSPVWAGGLFVADRKFWELGSDYD 240
QY 338 GLEIWSGEQYEISFKVMCGRMEDIPCRRVGHYRKVYPPKVPAGVSLARNLKRAEYV 397
DB 241 GLEIWSGEQYEISFKVMCGRMEDIPCRRVGHYRKVYPPKVPAGVSLARNLKRAEYV 300
QY 398 MDEYAEYIYORREYRHLISAGDVAVQKLRSSLNCKSPFKMFMTKIAMDLPKFPYVPEPPA 457
DB 301 MDEYAEYIYORREYRHLISAGDVAVQKLRSSLNCKSPFKMFMTKIAMDLPKFPYVPEPPA 360
QY 458 AAMGEIRNVGTGLCADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDQ 517
DB 361 AAMGEIRNVGTGLCADTKHGALGSPRLREGCVRGGEAAMNNQVFTFWREDIRGDDQ 420
QY 518 HTKFCFCDATISHTSPVTLVYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDSESDDHIFMNT 577
DB 421 HTKFCFCDATISHTSPVTLVYDCHSMKGNQJMKYRKDKTLVHPVSGSCMDSESDDHIFMNT 480
QY 578 CNPSSLTOQWLFHTNSTVLEKFN 603
DB 481 CNPSSLTOQWLFHTNSTVLEKFN 506

RESULT 4
US-09-795-926-29
Sequence 29, Application US/09795926
Patent No. 6555669
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 535
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 4; Length 535;
Best Local Similarity 94.6%; Pred. No. 1,3e-286; Indels 29; Gaps 1;
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

```

QY 98 MTDARVDAQYRENGNFIYVSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 157
DB 1 MTDARVDAQYRENGNFIYVSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 60
QY 158 WSSLLRTVSHVLRSPPELVAEIVLVDDFSDREHLKKPLEDMALPFSVAILTKKREG 217
DB 61 WSSLLRTVSHVLRSPPELVAEIVLVDDFSDREHLKKPLEDMALPFSVAILTKKREG 120
QY 218 IIRFMIGASVATGDTVITFLDSHCEANVMPLPPLLDRIARNRKTIIVCPMIDVINDHDPRE 277
DB 121 IIRFMIGASVATGDTVITFLDSHCEANVMPLPPLLDRIARNRKTIIVCPMIDVINDHDPRE 180
QY 278 TOAGDAMRGAFDWMYKRIPIPELQKADPSDFESPVMAGLFAVDRKFWELGSDYD 337
DB 181 TOAGDAMRGAFDWMYKRIPIPELQKADPSDFESPVMAGLFAVDRKFWELGSDYD 240
QY 338 GLEINGGEQYEISFK-----VMMCGRMEDIPCSR 368
DB 241 GLEINGGEQYEISFKHMLPRLVSNMPOAVFLPAPNMALQVMMCGRMEDIPCSR 300
QY 369 GHYRKVPYKVPAGVSLAHLKRVAVMMDEYAEYTORRPEYRHLSDGVAVOKKLR 428
DB 301 GHYRKVPYKVPAGVSLAHLKRVAVMMDEYAEYTORRPEYRHLSDGVAVOKKLR 360
QY 429 SLNCKSEKFMETKLANDLPKFPYVPEPPAAMGBIRNVGTGLCADIKGLGSPRLREG 488
DB 361 SLNCKSEKFMETKLANDLPKFPYVPEPPAAMGBIRNVGTGLCADIKGLGSPRLREG 420
QY 489 VRGRGEAAMNMVOFTTWMEDIRPGDPQHTKKCFPAISHTSPVLYDCHSKGNQJMK 548
DB 421 VRGRGEAAMNMVOFTTWMEDIRPGDPQHTKKCFPAISHTSPVLYDCHSKGNQJMK 480
QY 549 YRDKTLYHPVSGSCMDSESDHRIFMNTCNPSLSLTQOMLFEHTNSIVLEKFN 603
DB 481 YRDKTLYHPVSGSCMDSESDHRIFMNTCNPSLSLTQOMLFEHTNSIVLEKFN 535

RESULT 5
US-09-795-926-39
; Sequence 39, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abulin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 366
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-39

```

```

Query Match 57.3%; Score 1877; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 3.5e-193;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKKEKRLIQAVLVTAALVLLPNVGLMALYEROPDGPFGSGAAMVAPAAAGGSHSRQ 60
DB 1 MRKKEKRLIQAVLVTAALVLLPNVGLMALYEROPDGPFGSGAAMVAPAAAGGSHSRQ 60
QY 61 KTFPLDGGQKLDKMHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGNFIYVSDK 120
DB 61 KTFPLDGGQKLDKMHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGNFIYVSDK 120
QY 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEGSSLLRTVSHVLRSPPELVAEI 180
DB 121 ISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEGSSLLRTVSHVLRSPPELVAEI 180
QY 181 VLVDPSDREHLKKPLEDMALPFSVAILTKKREGIITRMLGASVATGDTVITFLDSHC 240
DB 181 VLVDPSDREHLKKPLEDMALPFSVAILTKKREGIITRMLGASVATGDTVITFLDSHC 240
QY 241 EAVNMPLPPLLDRIARNRKTIIVCPMIDVINDHDPFRYETOAGDAMRGAFDWMYKRIPI 300
DB 241 EAVNMPLPPLLDRIARNRKTIIVCPMIDVINDHDPFRYETOAGDAMRGAFDWMYKRIPI 300
QY 301 PELQKADPSDFESPVMAGLFAVDRKFWELGSDYDGLIENGGEQYEISFKV 353
DB 301 PELQKADPSDFESPVMAGLFAVDRKFWELGSDYDGLIENGGEQYEISFKV 353

RESULT 6
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abulin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

Query Match 41.8%; Score 1370; DB 4; Length 269;
Best Local Similarity 100.0%; Pred. No. 9.1e-139;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDAQYRENGNFIYVSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 157
DB 1 MTDARVDAQYRENGNFIYVSDKISLNRSLPDIRHPNCNKRYLETLPNTSIIIPPHNEG 60

```

Qy	158	WSSLRLRTVHSVNLNSPPELVAEI	VLVDDFSPREHKKLEDMALF	PSVRLIRTKRGGL	217
Db	61	WSSLRLRTVHSVNLNSPPELVAEI	VLVDDFSPREHKKLEDMALF	PSVRLIRTKRGGL	120
Qy	218	IRTRMLGASVATGVDITFLDSHC	EANVMWMLPPLLDRIARNRKTIV	CPMIDVI	277
Db	121	IRTRMLGASVATGVDITFLDSHC	EANVMWMLPPLLDRIARNRKTIV	CPMIDVI	180
Qy	278	TOAGDNRGATDWMYTKRIPRIPPEL	QKADPSDPFESVYMA	GGLFPAVDRKXFWELG	337
Db	181	TOAGDNRGATDWMYTKRIPRIPPEL	QKADPSDPFESVYMA	GGLFPAVDRKXFWELG	240
Qy	338	GLEIWGGEQVEYETISFKV	353		
Db	241	GLEIWGGEQVEYETISFKV	256		
RESULT 7					
US-09-795-926-35					
; Sequence 35. Application US/09795926					
; Patent No. 6555669					
; GENERAL INFORMATION:					
; APPLICANT: Donoho, Gregory					
; APPLICANT: Hilbun, Erin					
; APPLICANT: Turner, C. Alexander Jr.					
; APPLICANT: Friedrich, Glenn					
; APPLICANT: Abujin, Alejandro					
; APPLICANT: Zambrowicz, Brian					
; APPLICANT: Sands, Arthur T.					
; APPLICANT: Walke, D. Wade					
; APPLICANT: Wilganowski, Nathaniel L.					
; APPLICANT: Hu, Yi					
; APPLICANT: Kiele, James Alvin					
; APPLICANT: Potter, David George					
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND					
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME					
; FILE REFERENCE: LEX-0144-USA					
; CURRENT APPLICATION NUMBER: US/09/795, 926					
; CURRENT FILING DATE: 2001-02-28					
; PRIOR APPLICATION NUMBER: US 60/185,920					
; PRIOR FILING DATE: 2000-02-29					
; PRIOR APPLICATION NUMBER: US 60/186,558					
; PRIOR FILING DATE: 2000-03-02					
; PRIOR APPLICATION NUMBER: US 60/191,849					
; PRIOR FILING DATE: 2000-03-24					
; NUMBER OF SEQ ID NOS: 47					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 35					
; LENGTH: 321					
; TYPE: PRT					
; ORGANISM: homo sapiens					
US-09-795-926-35					
Query Match 40.0%; Score 1311; DB 4; Length 321;					
Best Local Similarity 99.2%; Pred. No. 2.8e-132;					
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;					

```

QY 241 EAVNYMLPPLIDR 253
Db 241 EAVNYMLPPLIGK 253

      RESULT 8
      US-08-967-508-19
      Sequence 19, Application US/08967508
      Patent No. 5910570
      GENERAL INFORMATION:
      APPLICANT: The Upjohn Company
      APPLICANT: For U.S. PURPOSES ONLY: Elhammer, Ake P. and Homa, Fred L.
      TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GALNAc:
      NUMBER OF SEQUENCES: 19
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
      ADDRESSEE: Property Legal Services
      STREET: 301 Henrietta Street
      CITY: Kalamazoo
      STATE: Michigan
      COUNTRY: USA
      ZIP: 49001
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent in Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/967,508
      FILING DATE:
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/602,830
      FILING DATE: 13 No. 5910570ember 1995
      ATTORNEY/AGENT INFORMATION:
      NAME: Darnley Jr., James D.
      REGISTRATION NUMBER: 33,673
      REFERENCE/DOCKET NUMBER: 4755-P CP
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 616-833-2210
      TELEFAX: 616-833-8897
      TELEX: 224401
      INFORMATION FOR SEQ ID NO: 19:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 517 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      US-08-967-508-19

Query Match      34.0%; Score 115; DB 2; Length 517;
Beet Local Similarity 43.6%; Pred. No. 8.3e-11;
Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16

QY 88 GNGEGGRP--YPMTDARVRVDQAYENGENIYVSDKISINSLPDIRHNCNCKXYLETLP 145
Db 15 GPGEKGRKVVI PKEDQEKMKEMFKINGNLMASEIMALNRSLLPVRLEGCKTKVYPDVLP 74

QY 146 NTSIIITIPHNKGSLLRTTHSVYNSRSPPELVAETIVLVDPSSDEHLKKPIEDTM-ALFP 204
Db 75 TTSSVIVFHNNAWSTLTLLRTTHSVYNSRSPRHMLESIIVLVDDASERDPFLRPLSESYVKLKV 134

QY 205 SVRIIRTKRGKGLRIRTMLGASVATGVITPFLDSHCENAVNMLPPLDIRIARNKRTIVCP 264
Db 135 PVHVIIRKGRGLIRARLKGAIVSKGVITITLDHCECTVGMLEPLARIGHDKRTIVCP 194

QY 265 MIDVIVHDHDFRYETQAG-DAMRGAFDWEMYYKRIPIPP--ELQKADPSDFESPVMVAG 320
Db 195 IIVIVSDTFEY--MAGSDMTVYGGPNMKLNFRWYVVPQREMDRRKGRDRTLVPRTPTVAGG 252

QY 321 LPAVDKKKFWELGQVDPGLIWTGGEQVETISKVMMCGRMMDICRSRGHIVYRKIVPVKV 380

```

```

Db 253 LFSDRDYFOBIGITVDAGMDIMGENSELEISFRIMQCGTLEIVTCSHVGHFRKATPYTE 312
QY 381 PAGVS--LARNLKRVAEVMWDEYAEYIYORREPRHLSAGDVAVQKLRSSLNCKSPKWF 438
Db 313 PGGTGQIINKNNRRLAEVWMDDEFKNFYIISPGVTKYVDGDISRLGLRHLQCRPSWY 372
QY 439 MTKIAMD--LPKFPYPPVEPPAAWGEIRNVGTGLCADTKHGALGSPRLBEGCYRGGEAA 496
Db 373 LENIYDPSQIPRHY-----FSLGEIRNVETWQCLDNNAKENEKXGIFNC--HGWG--- 421
QY 497 WNNMQUFTFWREDIRGDPQHTKKFCFPAISHTSPVTLYDCSHMKGNQMLKTRKDK--TL 555
Db 422 --GNQVPSYANKERITDD-----LCIDVSKLNGVPTMLKCHHLKGNQMLMEYDPVXLTL 473
QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTOQMLFEHTSTVLEKF 600
Db 474 QHVNNSQCJDKXATDEDSQVPSIRDCS-GSRSQOMLLR--NVTLPelf 517

```

RESULT 9

```

US-08-967-506-19
; Sequence 19, Application US/08967506
; Patent No. 6096512
; GENERAL INFORMATION:
; APPLICANT: The Upjohn Company
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc-
; TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company, Intellectual
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/967,506
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/602,830
; FILING DATE: 13 No. 6096512ember 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-833-2210
; TELEFAX: 616-833-8897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-967-506-19

```

Query Match 34.0%; Score 115; DB 3; Length 517;

Best Local Similarity 43.6%; Pred. No. 8.3e-11;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEGRP--YPMTDARVQAVRENGFNIVSDKISLNSLPDIRHFNCSKRYLETLP 145

```

Db 15 GPEMGKPVVLPKEDKMKEMKFNQFNLMSAMIALNRSJPDVLEGGCKTKVYPDNL 74
QY 146 NTSIIIPHNKESGLIRTVHSLVNSRPELVAEITLVDFSDREHLKKPLJEDYM--ALFP 204
Db 75 TTSVIVFHEANSTLIRTVHSLVNSRPRMIEIVLVDDASERDLKAPLESYVKKLY 134
QY 205 SVRIIRTKKRGELIRTRMLGASVATGVTITFLDSHCANVNMJPLLDRIARNKRTIVCP 264
Db 135 PVAHVMQSGSLIRARLGAASVKGQVITFDAAHECTVGMLEPLARIKHDRKTVVC 194
QY 265 MIVIVDHDPRRYTQNG--DAMGAFDMEVYKRIPIPP--ELQKADPSDPFSPVMAAG 320
Db 195 IIDVIGDDPEY--MAGSDMTYGGFNMKINFRWYPPVOREMRKRDRTLPAVPTMAAG 252
QY 321 LPAVDKRMFELGSGYRPGLEIMGEGYBISFKVMQMGGMEDPCSGVGHYIKVYRKY 380
Db 253 LFSDRDYFOEITVDAGMDIMGENSELEISFRIMQCGTLEIVTCSHVGHFRKATPYTE 312
QY 381 PAGVS--LARNLKRVAEVMWDEYAEYIYORREPRHLSAGDVAVQKLRSSLNCKSPKWF 438
Db 313 PGGTGQIINKNNRRLAEVWMDDEFKNFYIISPGVTKYVDGDISRLGLRHLQCRPSWY 372
QY 439 MTKIAMD--LPKFPYPPVEPPAAWGEIRNVGTGLCADTKHGALGSPRLBEGCYRGGEAA 496
Db 373 LENIYDPSQIPRHY-----FSLGEIRNVETWQCLDNNAKENEKXGIFNC--HGWG--- 421
QY 497 WNNMQUFTFWREDIRGDPQHTKKFCFPAISHTSPVTLYDCSHMKGNQMLKTRKDK--TL 555
Db 422 --GNQVPSYANKERITDD-----LCIDVSKLNGVPTMLKCHHLKGNQMLMEYDPVXLTL 473
QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTOQMLFEHTSTVLEKF 600
Db 474 QHVNNSQCJDKXATDEDSQVPSIRDCS-GSRSQOMLLR--NVTLPelf 517

```

RESULT 10

```

PCT-US94-02552-19
; Sequence 19, Application PC/TUS9402552
; GENERAL INFORMATION:
; APPLICANT: Elihammer, Ake P.
; TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc-
; TITLE OF INVENTION: Polypeptide, N-Acetylglucosaminyltransferase
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Upjohn Company, Corp. Intellectual
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/02552
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Darnley Jr., James D.
; REGISTRATION NUMBER: 33,673
; REFERENCE/DOCKET NUMBER: 4755.P CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 616-385-5210
; TELEFAX: 616-385-6897
; TELEX: 224401
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 517 amino acids
; TYPE: amino acid

```


ADDRESSEE: Property Legal Services
 STREET: 301 Henrietta Street
 CITY: Kalamazoo
 STATE: Michigan
 COUNTRY: USA
 ZIP: 49001
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Releasee #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/967,506
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/602,830
 FILING DATE: 13 No. 6096512ember 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Darnley Jr., James D.
 REGISTRATION NUMBER: 33,673
 REFERENCE/DOCKET NUMBER: 4755.P CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 616-833-2210
 TELEFAX: 616-833-8897
 TELEX: 224401
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 559 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-967-506-9

Query Match 34.0%; Score 1115; DB 3; Length 559;
 Best Local Similarity 43.6%; Pred. No. 9.5e-111;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

88 GNGEOGR--YPTDARVDOAYRENGFNIVYSDKISLNSRLPDIRHPNCNSRYLETLP 145
 57 GPEEMKPVVITPEDEKMKEMFKINOFNLMASEMIALNRLSPVLRBGGKTYVDPNLP 116
 146 NTSIIIPFNEGSSLLRTVSHVSNRSPELVAEIVLVDFSDREHLKKPLEDDYM-ALFP 204
 117 TTSVIVFHNEMASTLLRTVSHVSNRSPELMLEIVLVODASBRDPLKRLPSYVKLKY 156
 205 SVRIILTKKREGLIRRMIGASVATGDTITFLDSHCENAVNMLPPLIDRIARRKTIYCP 264
 177 PVHIVMEORSGLIRRLKGAASVSKQVITFLDAHCECTVGMLEPLARIKHDRKTVCP 236
 265 MIDVIDHDPFRYETOG-DAMRGAPDMEVYKRIPIP---ELQKADSPDPFSPVMAG 320
 237 IIDVIDSDTFEY--MAGSDMTYGGFWKLNFRMYPVPOEMDRKGRITLPVPTPMAGG 294
 321 LFAVDRKFMELGVDYDGLIENGGEQYEISFKVMCGRMEDIPCSRVGHVYRKVYKY 380
 295 LFSIDIDYFOEIGTYAGMDIWGGENLEISFRIMOGCGTLEIVTSGHVGHVFRKATPYTF 354
 381 PAGVS--LARNLKRVAEVMDEYAEIYQRRPEYRHLASGDVAVQKKLSSLNCSFKMF 438
 355 PGGTGQIINKNNRRLAEVWMDPEKMFYIISPGVTKVYGDIDISSRLGLRHKLQCRPFSSWY 414
 439 MTGIAMD--LPRKYPPEPAAAMGEIRVNGTGLCADTKHGLAGSLRLLEGCGVRGEAA 496
 415 LENIYVDSQIPRYH-----FSLGEIRVNETNOCDNARKENKVGATNC--HGNG--- 463
 497 WNNMQUFTPTWRDIDRPGDPOHTKFCFDAISHTSPVTLVYDCHSMKGNQMLKTRKDK-TL 555
 464 --GNGVFSYTNANKEIRDD-----ICLDVSKLNGVPTMLKCHHLKGNQJMEIDPYKTLI 515
 556 YHPSVSGSCMD-CSESDHRI-FNATCNPSSTLQOMLEFHTNSTYLEKPF 600

Db 516 QHVNNSQCLDKATDEDSQVPSIRDCS-GSRSQOMLR--NVTLPEIF 559

RESULT 13

PCT-US94-02552-9
 Sequence 9, Application PC/TUS9402552

GENERAL INFORMATION:
 APPLICANT: Elhammer, Ake P.

TITLE OF INVENTION: A Cloned DNA Encoding a UDP-GalNAc:
 Polypeptide, N-Acetylgalactosaminyltransferase

NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESSES:

ADDRESSEE: The Upjohn Company, Corp. Intellectual

STREET: 301 Henrietta Street

CITY: Kalamazoo

STATE: Michigan

COUNTRY: USA

ZIP: 49001

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Releasee #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US94/02552

FILING DATE:

CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:

NAME: Darnley Jr., James D.
 REGISTRATION NUMBER: 33,673

REFERENCE/DOCKET NUMBER: 4755.P CP
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 616-385-5210
 TELEFAX: 616-385-6897

TELEX: 224401
 INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:
 LENGTH: 559 amino acids

TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: protein

PCT-US94-02552-9

Query Match 34.0%; Score 1115; DB 5; Length 559;
 Best Local Similarity 43.6%; Pred. No. 9.5e-111;

Matches 230; Conservative 89; Mismatches 170; Indels 38; Gaps 16;

88 GNGEOGR--YPTDARVDOAYRENGFNIVYSDKISLNSRLPDIRHPNCNSRYLETLP 145
 57 GPEEMKPVVITPEDEKMKEMFKINOFNLMASEMIALNRLSPVLRBGGKTYVDPNLP 116
 146 NTSIIIPFNEGSSLLRTVSHVSNRSPELVAEIVLVDFSDREHLKKPLEDDYM-ALFP 204
 117 TTSVIVFHNEMASTLLRTVSHVSNRSPELMLEIVLVODASBRDPLKRLPSYVKLKY 156
 205 SVRIILTKKREGLIRRMIGASVATGDTITFLDSHCENAVNMLPPLIDRIARRKTIYCP 264
 177 PVHIVMEORSGLIRRLKGAASVSKQVITFLDAHCECTVGMLEPLARIKHDRKTVCP 236
 265 MIDVIDHDPFRYETOG-DAMRGAPDMEVYKRIPIP---ELQKADSPDPFSPVMAG 320
 237 IIDVIDSDTFEY--MAGSDMTYGGFWKLNFRMYPVPOEMDRKGRITLPVPTPMAGG 294
 321 LFAVDRKFMELGVDYDGLIENGGEQYEISFKVMCGRMEDIPCSRVGHVYRKVYKY 380
 295 LFSIDIDYFOEIGTYAGMDIWGGENLEISFRIMOGCGTLEIVTSGHVGHVFRKATPYTF 354
 381 PAGVS--LARNLKRVAEVMDEYAEIYQRRPEYRHLASGDVAVQKKLSSLNCSFKMF 438
 355 PGGTGQIINKNNRRLAEVWMDPEKMFYIISPGVTKVYGDIDISSRLGLRHKLQCRPFSSWY 414

Db 520 TIFPHSGTCLISAYRTEGRPDVQWRTCDALDKXQJWSFE 559

RESULT 16

US-09-795-926-33

Sequence 33, Application US/09795926
Patent No. 6555669
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedlich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 33
LENGTH: 240
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-33

Query Match 30.2%; Score 990; DB 4; Length 240;

Best Local Similarity 97.9%; Pred. No. 6.8e-98;
Matches 189; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MRKREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAOGSHSRK 60
Db 1 MRKREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAOGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDARQVNGEGRPYMTDAERVDQAYRENGFNIVSDK 120
Db 61 KTFPLDGGQKLKDMHDKAIRDARQVNGEGRPYMTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPFHNMGSSLLRTVHSVLRSPPELVAEI 180
Db 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPFHNMGSSLLRTVHSVLRSPPELVAEI 180
QY 181 VLVDFFSDRRLK 193
Db 181 VLVDFFSDRGRIR 193

RESULT 17

US-09-795-926-37

Sequence 37, Application US/09795926

Patent No. 6555669
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedlich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.

APPLICANT: Walke, D. Wade
APPLICANT: Wilganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 37
LENGTH: 209
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-37

Query Match

30.1%; Score 988; DB 4; Length 209;
Best Local Similarity 100.0%; Pred. No. 8.8e-98;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAOGSHSRK 60
Db 1 MRKREKRLQAVLVLAALVLPVGLMALYRERQDPGTGGSGAAYAPAAOGSHSRK 60
QY 61 KTFPLDGGQKLKDMHDKAIRDARQVNGEGRPYMTDAERVDQAYRENGFNIVSDK 120
Db 61 KTFPLDGGQKLKDMHDKAIRDARQVNGEGRPYMTDAERVDQAYRENGFNIVSDK 120
QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPFHNMGSSLLRTVHSVLRSPPELVAEI 180
Db 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPFHNMGSSLLRTVHSVLRSPPELVAEI 180
QY 181 VLVDFFSDR 189
Db 181 VLVDFFSDR 189

RESULT 18
US-10-140-002-196
Sequence 196, Application US/10140002
Patent No. 6725730
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerltsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Thomas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C59
CURRENT APPLICATION NUMBER: US/10/140,002
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper

NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 196
 LENGTH: 552
 TYPE: PRT
 ORGANISM: Homo Sapien
 US-10-140-002-196

Query Match 29.6%; Score 971; DB 4; Length 552;

Best Local Similarity 37.2%; Pred. No. 3,1e-95; Mismatches 199; Indels 98; Gaps 19;

Matches 229; Conservative 90; Mismatches 199; Indels 98; Gaps 19;

QY 1 MRREKELQAVLVLAALVLLPVGVMALYRERQ-PDG---TPGSGAAVAPAAQGS 55
 1 MRRLTRLVLPVFEVLTITVL---FFWYTRKLEVTGPEVQPKRSDA----- 47
 DB 56 HSROKTFPLGDDQCKLQMDH-----KEAIRDAQVNGEGGPRYPMTDAERVDQAY 108
 48 -----DWDLDMDQFDERRYLNAKKRVG-----DDPY 74
 QY 109 RENGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHSGMSLLRTVHSV 166
 75 KLVAFNRESERISSNPAIPDTRHLRCTLVYCTDLPPTSIIIFHNANSTLRTIRSV 134
 QY 169 LNSPPELVAEIYLVDPFS-DREHLKKPLEDYALPFSVRLTRTKKEGLITRMGLASV 227
 135 LNRTPHLIRHLLVDPFSDNDPDDCKLIK-----LPKVKCLRNERRGQLVRSIRGADI 189
 QY 228 ATGCVITFLDSHCANVNLPLLDRIARNKTIYVCPMIDIVDHDPRYETQAGDAMRGA 287
 190 AOGTTLTFLDSHCENRDWLQPLHRYVEDYTRVYCEVIDIINDFTYIESASE-LRG 248
 QY 288 FDMEMYKRIPIPELOKA---DPSDPESPVMAGLFAVDRKFMELGVDPELITWG 344
 249 FDSLHQQWEOLSE-QARRLDPEPTPTITLNGSLFVIDKAFDYLKGMMDMIDWG 307
 DB 345 EOYISFKVMCGRMEDIPCSRGIHYRYKVPYKVPAG-VSLARLKVAAVWMDYA 402
 308 ENFISIRVWVGSGSLFVPCSRGVHFRKHPVFPDGNANTYIKTKRAEAWMDYK 367
 QY 403 EYIYORPEYRHLSAGDVAQVKLRSSLNCSFFWMTKIAMDLPKFYPPVEPP--AAA 459
 368 QYVVAARPFALERPFVGVESHLDLRKMLRQSPFKVYLENI-----YPELSIPKESSIQ 420
 DB 460 WGEIRNVTGTCACATKIGALSPRLLEGCVRGGEAAMNNQVFTFTREDIRGCDQHT 519
 421 KGNRQROKCLSESQONNOETPNLKLSPCAVKGEDA-KSQVAAFTYTOQIL-----Q 472
 QY 520 KKFCEDAIS--HTSPVTLVDCHSMKGNOLWKYRKDXTLYHPVSGSCMDC-----SSDH 571
 473 BECLSVITLFPKAPVVLVLCNKGDDROQT-KTGSHEIHAHLCLDITMFGDGTENGK 531
 QY 572 RIFMNTCPSSLTQW 587
 DB 532 EIVNPPCESSIMSQHW 547

RESULT 19

US-09-270-767-33733 Application US/09270767

Sequence 33733, Patent No. 6703491

GENERAL INFORMATION:

APPLICANT: Homburger et al.

TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster

FILE REFERENCE: File Reference: 7326-094

CURRENT APPLICATION NUMBER: US/09270,767

NUMBER OF SEQ ID NOS: 62517

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 33733

LENGTH: 302

TYPE: PRT

ORGANISM: Drosophila melanogaster

US-09-270-767-33733

Query Match 28.7%; Score 941; DB 4; Length 302;

Best Local Similarity 59.7%; Pred. No. 1.9e-92; Mismatches 44; Mismatches 70; Indels 6; Gaps 4;

Matches 178; Conservative 44; Mismatches 70; Indels 6; Gaps 4;

QY 80 IRDAQVNGEGGPRYPMTDAERVDQAYR---ENGFNIVSDKISLNSLPDIRHPN 136
 2 MENDARVGLGEGGKASTLDDSDQLERKMSLENGFNMLLSISVNRSPVDIRHPLCR 61
 DB 137 SKRYLETPTSIIIPFNHSGMSLLRTVSVNRPPELVAEIYLVDPFSDEHLKPL 196
 62 KKEVYAKLPVSYIIIFVNEYLSVLMRSVSLNRPPELMKXIIIVDHSDEHYLQEL 121
 QY 197 EDYMA-LFESVRLTRTKKEGLITRMGLASVATGVITFLDSHCANVNLPLLDRIA 255
 122 EYIAEHFKVVRVRLPRRTGLIGAPAGARNAETVLFDSHVEANTYMLFPLLEPIA 181
 DB 256 RNRKTIYCPMIDIVDHDPRYETQAGDAMRGAFDMEMYKRIPIPELOKADSDPESP 315
 182 LNRRTAVCPEDIVDHTNFFHYRAQ-DEGARAGATWFEFYKRLPLPEDLK-HPADFPKSP 239
 QY 316 WAGGLFAVDRKFMELGVDPELITWGSGEYISFKVMCGRMEDIPCSRGIHYR 373
 240 IMAGGLFAISKEPFWEIGYDEGLDIWGEQYELSKRIWCGSEMYDAPCSRGIHYR 297

RESULT 20

US-09-376-856-2

Sequence 2, Application US/09376856

Patent No. 623510

GENERAL INFORMATION:

APPLICANT: KIMAR, SANJAY

APPLICANT: VAN HORN, MARION M.

TITLE OF INVENTION: LARK, MICHAEL

FILE REFERENCE: GP-70638

CURRENT APPLICATION NUMBER: US/09376,856

NUMBER OF SEQ ID NOS: 2

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2

LENGTH: 639

TYPE: PRT

ORGANISM: HOMO SAPIENS

US-09-376-856-2

Query Match 28.2%; Score 926; DB 3; Length 639;

Best Local Similarity 35.5%; Pred. No. 2.8e-90; Mismatches 216; Indels 80; Gaps 17;

Matches 216; Conservative 96; Mismatches 216; Indels 80; Gaps 17;

QY 4 KERKLLQAVLVLAALVLLPVGVMALYRERQDGTTPGSGAAVAPAAQGSRSRQKTF 63
 91 REDQLVAVLPOA-----RNDQOGRRGGSYRLIKOPRRQ----- 126
 DB 64 FLDDGQCKLQMDHKEAIRDAQVNGEGGPRYPMT---DAERVDQAYENGFNIVSD 119
 127 -----DKEAFKRWGADDEGESEELIPFSLDPGLQEA-----LSA 165
 QY 120 KISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNHSGMSLLRTVSVNRPPELVAE 179
 166 RIFLQRLPEVVRHPLCQHPQDSLPFASVITLGFHDEAMGTLRTVSHLIDTVPRAPLKE 225
 DB 180 IYLVDPFSDEHLKPLLEDMALPFSVRLTRTKKEGLITRMGLASVATGVITFLDSH 239
 226 IIVVDLSQOGQKSLASEVVARLEGVKLRNSKRGALRAARMLGATRAVDLVFWDAAH 285
 QY 240 CEANVNMPLPLDIRIARNKTIYVCPMIDIVDHDPRYETQAGDAMRGAFDMEMYKRIPI 299
 286 CECHPGWLEPLSRIGADRVRVSPYIVDMKTFQY-YPSKLOQGVLDKIDPFHKEPL 344
 DB 300 PPELOKA-DPSDPESPVMAGLFAVDRKFMELGVDPELITWGSGEYISFKVMCG 357
 345 PEHVRKALQSPISPIRSPPVVGSEVAMDRHYFQNTGAYDLSMLRGGENLELSFKAWLCG 404

```

Oy 358 GRMEIDPCSRVQHIYKRYKVPYKVPAGVSLAARNLKRAEYWMDBYAYIYORRREYHLSA 417
Db 405 GSVELIFPCSRVGHIIYONQDSHSPBDEAETILRRKRVRLAETWLSGFKETPKYHSEASLSK 464
Oy 418 G---DVAVOKKJLRSSLNCKSPKFMPTKIAMDLPEKYRPEVEPPAAANGELRNVTGLCAD- 473
Db 465 AEKRPOMERLQIQRRLGCRTPFMWFLANV---YELTPS--EPREPSFGKLNITLGLCADC 520
Oy 474 -TKHAGLSPLLEGGVGRGEAAMNMVVFPTTMEDELRPGPQHTKKCFPAISHTSP 532
Db 521 QAEIGILGCPMVLAPCSDSR-----QOQYLQTSKKEIHFGSPQH--LCFAN--RQEQ 569
Oy 533 VTLYPDCHSNKQ---NQLMKYRKDKTLVHPVSGCNDG--SESDBHIFMNTCNPSLSLTOQ 586
Db 570 VILQNC-TEEGGLAIHQOHMDPQBNMGWIVHLSGCKMEAVQENNKDLYLRPCD-GRARQ 627
Oy 587 WLEEHNTNS 594
Db 628 WRFQIINA 635

```

```

RESULT 21 488-2
US-09-347-488-2
: Sequence 2, Application US/09347488
: Patent No. 6239266
: GENERAL INFORMATION:
: APPLICANT: Munroe, David
: APPLICANT: Prihall, Ingrid
: TITLE OF INVENTION: ZAP-3 TUMOR ASSOCIATED GENES AND THEIR
: TITLE OF INVENTION: US-8
: FILE REFERENCE: GEN-2PRV
: CURRENT APPLICATION NUMBER: US/09/347,488
: CURRENT FILING DATE: 1999-07-02
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 638
: TYPE: PRT
: ORGANISM: H. sapiens
US-09-347-488-2

```

Query Match	28.1%;	Score 920;	DB 3;	Length 638;
Best Local Similarity	35.4%;	Pred. No. 1.2e-89;		
Matches 215; Conservative	96;	Mismatches 217;	Indels 80;	Gaps 17;

```

Oy      4 KEKRLLOVALVLAALVLLPNTGLMALYERODGTPGSGAIVAPAAAGGSHSRQKTF 63
Db      91 REDOLLVAVLPOA-----RRNOSQGRGGSYRLIKQPRR----- 126

Oy      64 FLGDGQKLDWMDKEAIRDAQRVNGEGORPYPMT----DARVOOAYRENGFNIVSD 119
Db      127 -----DKEAPKRDWGDGEDGEVSEEBELTPFLDPRGLGEA-----LSA 165

Oy      120 KLSLNRSLPDIRHPNCNSKRYETLPNTSIILIPFNEGSSLLRTYHSLVNRSPPELVAE 179
Db      166 RPLORALPEVHHPICTOQHPDPSIFTAASVILCFHBEAWSTLLRTYHSILDTVPRAFLKE 225

Oy      180 IYLVDFSPRHLKKRPLEDYMALFESVRLIRTKKRGGLRTWMLGASVATGVDYITLDSH 229
Db      226 IILVDDLSQGGQKSLSEYVARLSEGVKLLRSNKRSLAIPARMLGATRAAGDVLVPMDAH 285

Oy      240 CEANVNWLPPLLDRIARNRKTIVCPMIIVILDHDFRETQOAGAMGAFPMEWYKRIPI 299
Db      286 CECHGGMLEPLLSRIAGDRSRVSPYIVIDIMKTFPY-YP SKLDQGVLDKLDPFHMEPL 344

Oy      300 PPELOXA--DPSDPFESPVMAGGLFAVDKRWEMELGSGYDPGLEIWGEGOYEISFKVMWCG 357
Db      345 PEHVAKALQSPISPIRSPVVGAVMDRHHYFQNTGAYDLSMLRGGENLELSFKWMLCG 404

Oy      358 GRMEIDPCSRVGHYIRKTYPIKYVAPAGVSLARNLKRVAEVMMDYAEYIYQRRPEYHLSA 417
Db      405 GSVELILPCSRVGHYIYQNDOSHPLDQEAURLRNKVRILAEIWLGSFKTTPYHGSHSEARSLSK 464

```

```

QY      418  G---DVAVKOKRSSLNCKSPFMPKTXLAMDLPKRYPPVEBPAAWMEIRNVGGLCAD  473
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      465  AEPKDCMERKIQORRLCCTTFMFLAVN---YPELYPS-EFPSPFGKLNHTGLGLCAD  520
QY      474  -TKHGLGSPRLREGCVRGEGEAAVNNQVTFPTWREDIRPGDBQHTKKECFDAISHTSP  532
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      521  QAGSDILGCEWVLAPCSDSR-----QQQYQHTSRKKEIHGSPQH---LCFANV--RQEQ  569
QY      533  VTIYDCHSKMG---NQLMKRYKDKTLVHPVSGSCMDC--SESQHRIFMNTCNPSLSITQ  586
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      570  VILQNC-TEEGGALIHQOHWDFOENGMIIVHILSGKMEAVVQENNKDLYLRPCD-GKARQQ  627
QY      587  WLPEHTNS  594
        |||::|||::
Db      628  WRFDQJNA  635

```

RESULT 22
US-08-648-298-2
: Sequence 2, Application US/08648298

APPLICANT: Henrik Clausen
 APPLICANT: Eric Paul Bennett
 TITLE OF INVENTION: UDP-N-acetyl-alpha-D-galactosamine polypeptide
 TITLE OF INVENTION: N-acetylglucosaminyltransferase GalNAc-T3
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Darby & Darby PC
 STREET: 805 Third Avenue
 CITY: New York
 STATE: NY
 ZIP: 10022
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/648,298
 FILING DATE: 15-JUN-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Green, Reza
 REGISTRATION NUMBER: 38,475
 REFERENCE/DOCKET NUMBER: 4035/0865
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212527700
 TELEFAX: 2127536237
 TELEX: 236687
 INFORMATION FOR SEQ. ID NO. 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 633 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Submaxillary gland

Query Match	Score	DB 2	Length
27.9%	916	633	

Best Local Similarity 37.6%; Pred. No. 3.3e-89;
Matches 303; Conserved 100; Mismatches 177; Indels 59; Gaps 18

[illegible]

90 GEOGRPYMTDAERVDQAYRENG-----FNIVYSDKISLNRSL-PDIRHPNCNSKRY--L 141

Db 122 GASGKAEKTTNI.SVEEOKERKERGAEAKHCFNAPASDRIST.HRDL.GPDTBPPECTIEOKFKXC 181

[illegible]

142 ETLPNTSIIIPFHNEGSSLRTVHSLNRSPPELVAEIVLVDDESDREHLKKPLEDYMA 201

db 182 PPLPTSVIIVFHNEAWSTLLRIVHSVL YSSPAILLKEIILVDDASVDEYLDHDKLDEYVK 241

QY 202 LPSVRLTRKREGI.LRTMLGASVATGDIVITFLDSHCENAVNMLPPLDRIRANKTI 264
 242 QFSIVKIVROBERGGLTARLGLATVATETITFLDHCFCYGMLEPLARIENTAV 301
 QY 262 VCPMIDVIDHDDFRYTOA---GDAMGAPDMENY---KRIPIPELOKADPSDPFSPV 316
 302 VSPDIASIDINTFENKPSPYGSNNHGNFDMSLSPGMSL.PDHEKQRNDQETPIKTP 361
 QY 317 MAGGLFAVDKRMFELGCDPLGELIWGGEQYEISFKYMWCGRRMEDI.PCSRVGHIYKTY 376
 362 FAGGLFISKEFYKFEIYSYDEMEIWMGBENIEMSFVWQCGGLEIMPCSVVGHVFSKS 421
 QY 377 PYKYPAGVS-LARLKKVAEYVMDYAEYIYORPE---YRHLSAGDVAOKLBSLSLN 431
 422 PHSPFKGTQVYARQVRLAEVMDYKIEIFRRRTDAKTYKQAFGLSKREIFKIRLR 481
 QY 432 CKSEKFWMTKIAMDLPKFYPPVEPPAAMGEIRNVGTGLCADT-KHGLASPLRLBGCVR 490
 482 CKNTWTYLNMI---YPEVYVDLNPVLS-GYIKSVGQPLCLDVBENNOCGKPLIMYTC-H 536
 QY 491 GREGAANNNOVFTFTREDIRPDPOHT--KKCFDAI-----SHTSPVTLY 536
 537 GLG-----GNQYFYSQHEIR-----HNIOKELCIHAAGLVOLKACTYKGHKTIVT-- 584
 QY 537 DCHSMKGNOLMKYRKDKTLVHVSQCMDCSESDRI.FMNTCNPSLSLQCKLFEHTN 593
 585 -----GEOIWEIOKDQILNPLKMLCS-ANGHPSLV-SCNPSDPLQKWLISQND 633
 Db

RESULT 23

US-09-795-926-23
 Sequence 23, Application US/09795926
 Patent No. 6553669
 GENERAL INFORMATION:

APPLICANT: Donoho, Gregory
 APPLICANT: Hilbun, Erin
 APPLICANT: Turner, C. Alexander Jr.
 APPLICANT: Friedrich, Glenn
 APPLICANT: Abujin, Alejandro
 APPLICANT: Zambrowicz, Brian
 APPLICANT: Sands, Arthur T.
 APPLICANT: Walke, D. Wade
 APPLICANT: Wilganowski, Nathaniel L.
 APPLICANT: Hu, Yi
 APPLICANT: Kieke, James Alvin
 APPLICANT: Poter, David George
 TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 FILE REFERENCE: POLYNUCLEOTIDES ENCODING THE SAME
 CURRENT APPLICATION NUMBER: US/09/795, 926
 CURRENT FILING DATE: 2001-02-28
 PRIOR APPLICATION NUMBER: US 60/185,920
 PRIOR FILING DATE: 2000-02-29
 PRIOR APPLICATION NUMBER: US 60/186,558
 PRIOR FILING DATE: 2000-03-02
 PRIOR APPLICATION NUMBER: US 60/191,849
 NUMBER OF SEQ ID NOS: 47
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 23
 LENGTH: 224
 TYPE: PRT
 ORGANISM: homo sapiens
 US-09-795-926-23

Query Match 24.5% Score 804; DB 4; Length 224;
 Best Local Similarity 98.7%; Pred. No. 6.7e-76;

Matches 154; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 98 MTDARVDOAVRENGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNEG 157
 1 MTDARVDOAVRENGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNEG 60
 Db

QY 158 WSSLRTVSHVLSNRSPPELVAEIVLVDDFSDREHLKKPLEDYALPESVRLTRKREGI 217
 61 WSSLRTVSHVLSNRSPPELVAEIVLVDDFSDREHLKKPLEDYALPESVRLTRKREGI 120
 Db
 QY 218 IRTMLGASVATGDIVITFLDSHCENAVNMLPPLDR 253
 121 IRTMLGASVATGDIVITFLDSHCENAVNMLPPLDR 156
 Db

RESULT 24

US-09-973-457-2
 Sequence 2, Application US/0973457
 Patent No. 6703230
 GENERAL INFORMATION:

APPLICANT: Kapeller, Libermann, Rosana
 TITLE OF INVENTION: 47174, A NOVEL HUMAN GLYCOSYLTRANSFERASE
 FILE REFERENCE: 10448-099001
 CURRENT APPLICATION NUMBER: US/09/973,457
 CURRENT FILING DATE: 2001-10-09
 PRIOR APPLICATION NUMBER: 60/238,849
 PRIOR FILING DATE: 2000-10-06
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 2
 LENGTH: 603
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-973-457-2

Query Match 22.7% Score 744; DB 4; Length 603;
 Best Local Similarity 32.3%; Pred. No. 1e-70;

Matches 202; Conservative 86; Mismatches 230; Indels 108; Gaps 23;
 QY 3 RKEKRLIOAVLALVLAALVLLPNVGLMALYEROPDGTGGGGAAPAG-QGSHSROK 61
 5 RKIRTLTLVNIIVFVGIV-----LPSVGRLO-----GMSGLVNIIVSGDRVRSNHA 53
 Db

QY 62 TPLFGDGQKLDN-HDKCAIRDAQRY-----GNGEQGRPYM-TDAERYDOA 107
 54 VGTLDGREALQRLDHLHEEVYVNLGLAKPIGLVGGGGLGGGLAATLRDGOEABGK 113
 Db
 QY 108 YRNGFNIVSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPFNEGMSLSLRTVSH 167
 114 YEEGYNAQUSDSISLDRSIPDYPRKCRMSYAQDPQVSVFIFVNEALSVILNSVSH 173
 Db
 QY 168 VLNRSPPELVAEIVLVDDFSDREHLKKPLEDYV-ALFPS-VRIARTKREGI.LRTMLGA 225
 174 VVHNTBSQLKEIYLVDDNSDNTLKNFLDQYNNKYPGLVKTIVRSRRRGLIRARLQGW 233
 Db
 QY 226 SVATGDIVITFLDSHCENAVNMLPPLDRIRANKTIYCPMIDVIDHDDFRYTOAGDAMR 285
 234 KAATAVAVGPFDAHVEFNTGMAPALSRIREDRRRIYLPADINIKYSTFEVQ-QYANAAN 292
 Db
 QY 286 GADPWENYTKRIPIPEP-LQKADPSDPFSPVMAAGL.PAVDKMFMFELGCDPLGELI 343
 293 G-VNMGIMCMYI-IPQDWLDRGDESAPIPTPMIGCSFVADREYFGDGLDLPGEVYG 350
 Db
 QY 344 GEOYEISFKYMWCGRRMEDI.PCSRVGHIYKTYPPYKYPAGVSLARNIKRVAEYVMDYAE 403
 351 GENVELGMRYWQCGSNEVLPCSRVAHIEKTRPYNNDDIYAKRANALAAEYVMDYDFKS 410
 Db
 QY 404 YIYQR--RPEYHLSAGDVAVOKLRSSLNCSFKWFMWTKIAMDLPKFYPP--VEPPAA 458
 411 HHVYANNI.PMSNPGVDGVSERLALRQRLKCRSFKYLENV-----YPERKVVYNTL 463
 Db
 QY 459 AMGEIRV-VGTGLCADTKHGLASPLRLBGCYVARGGAANNNOVFTFTREDIRPDPO 517
 464 TYGVRNSKASAVCLD--QGA-----ED--GD-- 486
 Db
 QY 518 HTKFCFDAISHTSPVTLVYDCHSMKGNOLMKYRKD-----KTLVHVSQCMDCSE 568

Db 487 -----RAILYPCHGM-SSQVRYNSADGULLQPLGSLTAFLPDS-KCLVDVG 530
 QY 569 SDHRIFMNTCNPSLSLTQOMLFEBHTNS 594
 Db 531 TGRMPTLKCEBDVAPRTQRLMDPTQS 556

RESULT 25
 US-09-270-767-45334
 ; Sequence 45334, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 45334
 ; LENGTH: 289
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-45334

Query Match 20.6%; Score 675; DB 4; Length 289;
 Best Local Similarity 50.2%; Pred. No. 8.2e-64;
 Matches 131; Conservative 45; Mismatches 77; Indels 8; Gaps 4;

QY 111 NGFNIVYSDKISLNSRLSPDIRHPNCSKRYLETLPNTSIIPFHEGMSLLRTVHSLN 170
 Db 34 NRFNGEASDALPSNRDIPDRNPMCRKTKYREDLPETSVITTFHEARSTLIRTVSLN 93
 QY 171 RSPPELVAEIYLVDSGDRHKKPLDYNALPFSVRIKTKREGILRTRMGASVATG 230
 Db 94 RSPPELVAEIYLVDSGDRHKKPLDYNALPFSVRIKTKREGILRTRMGASVATG 149
 QY 231 DVTFELSHCEANVMNPLPDLRIARNRKTIYCPMIVIDHDDRYETQAGDAMRGAFDW 290
 Db 150 SVLFELDSHVECNMMLPRLERREDPTRYVCPVIVISMDNQYIGASAD-LRGGFDW 208
 QY 291 EMYKRIPIPE--LQKADPSDFESPVMAGLFAVDKRFMELGSDYDGLTIGGEQY 347
 Db 209 NLIFKWEIYLSPESEAMHNDPTTIRTPMIAAGLFLVDKAYFNGLGKYDKMMDWGSENL 268
 QY 348 EISFKVMCGGRMEDIPCSR 368
 Db 269 EISFRVQCGGSLIIPCSR 289

RESULT 26
 US-09-270-767-46411
 ; Sequence 46411, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 46411
 ; LENGTH: 238
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-46411

Query Match 17.1%; Score 561; DB 4; Length 238;
 Best Local Similarity 50.7%; Pred. No. 1.2e-51;
 Matches 113; Conservative 35; Mismatches 69; Indels 6; Gaps 4;

QY 71 LKDMHDKAIRRDQGVNGEGRP--YPMTDARVDAQVRENGFNIVYSDKISLNSRLP 128
 Db 19 LKDMFLAPSVQEAQK--PQEMCKPVKIPADMDLMKEKFKENQFLNASDMSLSRLT 76
 QY 129 DIRHPNCSKRYLETLPNTSIIPFHEGMSLLRTVHSLNRPPELVAEIYLVDSFD 188
 Db 77 DVRECCRRGHVASKPPTTSIVIFPHEAMWTLTLRTVWSVINSPPALLKEIILVDAS 136
 QY 189 REHLKKPLDYMALFP-SVRIILRTKREGILRTRMGASVATGVTTFPLDSHCEANVMN 247
 Db 137 RDLFGQLEBYVAKLPVKTFLVTERKSGILRRLGAERHSGEVIITFLDACECTEGWL 196
 QY 248 PLLDRIARNRKTIYCPMIVIDHDDRYETQAGDAMRGAFDW 290
 Db 197 EPLAIIVQNRKTIYCPMIVISDEFEYIT-ASDSTWGFNM 238

RESULT 27
 US-09-270-767-43543
 ; Sequence 43543, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: Patentln Ver. 2.0
 ; SEQ ID NO 43543
 ; LENGTH: 366
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; FEATURE:
 ; OTHER INFORMATION: Xaa means any amino acid
 US-09-270-767-43543

Query Match 15.4%; Score 503.5; DB 4; Length 366;
 Best Local Similarity 44.0%; Pred. No. 3.7e-45;
 Matches 120; Conservative 41; Mismatches 99; Indels 13; Gaps 6;

QY 14 LVLAALVLEPNVGL--MALYRROPDGTGGSGAAVAPAGGSHSRKKTFFLDGOK 70
 Db 97 LALCLVLPVLPVLANMSDHNKRVQEAHYTRFG--PKFAHQRLBGRPREVKLVGD-- 152
 QY 71 LKDMHDKAIRRDQGVNGEGRPYPMTDARV-DQAVRENGFNIVYSDKISLNSRLP 129
 Db 153 LGNFERKDV---KPRSGGEGENGHSLSPDKKMSDASEWEGMNIACSDBSIMHRSVVD 208
 QY 130 IRHPNCSKRYLETLPNTSIIPFHEGMSLLRTVHSLNRPPELVAEIYLVDSFSDR 189
 Db 209 TRLEBGRHNDYPPDLRTISVIIFHNEGFSVLMKRTVSHYDINSPTMLHEIILVDPSDK 268
 QY 190 EHLKKPLDYMALFP-SVRIILRTKREGILRTRMGASVATGVTTFPLDSHCEANVMN 248
 Db 269 ENLRSGLDDEVYLFQKGLVYVIRKEREGLIRTSRGAMEATGEVIVFLDAHGEVNTNMLP 328
 QY 249 PLLDRIARNRKTIYCPMIVIDHDDRYETQAG 281
 Db 329 PLLAPIYRDRVTWTVPIIDGIDHKNFEYRPVVG 361

RESULT 28
 US-09-795-926-21
 ; Sequence 21, Application US/09795926
 ; Patent No. 6555669
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Erin
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedlich, Glenn
 ; APPLICANT: Abuhin, Alejandro

```

: APPLICANT: Zambrowicz, Brian
: APPLICANT: Sands, Arthur T.
: APPLICANT: Walke, D. Wade
: APPLICANT: Wilganowski, Nathaniel L.
: APPLICANT: Hu, Yi
: APPLICANT: Kleke, James Alvin
: APPLICANT: Potter, David George
: TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
: FILE REFERENCE: LEX-0144-USA
: CURRENT APPLICATION NUMBER: US/09/795,926
: PRIOR FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 60/186,920
: PRIOR FILING DATE: 2000-02-29
: PRIOR APPLICATION NUMBER: US 60/186,558
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: US 60/191,849
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 21
: LENGTH: 143
: TYPE: PRT
: ORGANISM: homo sapiens
: US-09-795-926-21

Query Match      14.7%; Score 483; DB 4; Length 143;
Best Local Similarity 95.8%; Pred. No. 1.2e-43;
Matches 92; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      98 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIHRPNCNSKRYLETLPNTSIIIPFNEG 157
DB      1 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIHRPNCNSKRYLETLPNTSIIIPFNEG 60

QY      158 WSSLIRTVSHSVLNRSPELVVAEIVLVDDFSDRHLK 193
DB      61 WSSLIRTVSHSVLNRSPELVVAEIVLVDDFSDRGRIR 96

RESULT 29
: US-09-795-926-25
: Sequence 25, Application US/09795926
: Patent No. 6555669
: GENERAL INFORMATION:
: APPLICANT: Donoho, Gregory
: APPLICANT: Hilbun, Erin
: APPLICANT: Turner, C. Alexander Jr.
: APPLICANT: Friedlich, Glenn
: APPLICANT: Abuin, Alejandro
: APPLICANT: Zambrowicz, Brian
: APPLICANT: Sands, Arthur T.
: APPLICANT: Walke, D. Wade
: APPLICANT: Wilganowski, Nathaniel L.
: APPLICANT: Hu, Yi
: APPLICANT: Kleke, James Alvin
: APPLICANT: Potter, David George
: TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
: FILE REFERENCE: LEX-0144-USA
: CURRENT APPLICATION NUMBER: US/09/795,926
: PRIOR FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 60/185,920
: PRIOR FILING DATE: 2000-02-29
: PRIOR APPLICATION NUMBER: US 60/186,558
: PRIOR FILING DATE: 2000-03-02
: PRIOR APPLICATION NUMBER: US 60/191,849
: PRIOR FILING DATE: 2000-03-24
: NUMBER OF SEQ ID NOS: 47
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 25
: LENGTH: 112
: TYPE: PRT
: ORGANISM: homo sapiens
```

```

US-09-795-926-25

Query Match      14.7%; Score 481; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-43;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      98 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIHRPNCNSKRYLETLPNTSIIIPFNEG 157
DB      1 MTDARVDQAYRNGFNIIYSDKISLNRSLPDIHRPNCNSKRYLETLPNTSIIIPFNEG 60

QY      158 WSSLIRTVSHSVLNRSPELVVAEIVLVDDFSDR 189
DB      61 WSSLIRTVSHSVLNRSPELVVAEIVLVDDFSDR 92

RESULT 30
: US-09-973-457-6
: Sequence 6, Application US/09973457
: Patent No. 6703230
: GENERAL INFORMATION:
: APPLICANT: Kapeller-Libermann, Rosana
: TITLE OF INVENTION: 47174, A NOVEL HUMAN GLYCOSYLTRANSFERASE
: FILE REFERENCE: 10448-099001
: CURRENT APPLICATION NUMBER: US/09/973,457
: PRIOR FILING DATE: 2001-10-09
: PRIOR APPLICATION NUMBER: 60/238,849
: PRIOR FILING DATE: 2000-10-06
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 149
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus sequence
: US-09-973-457-6

Query Match      13.7%; Score 450.5; DB 4; Length 149;
Best Local Similarity 53.1%; Pred. No. 4.3e-40;
Matches 78; Conservative 28; Mismatches 36; Indels 5; Gaps 2;

QY      305 KADSPDPFESPVNAGGLFAVDRKMFWEIGGYDGLIWMGGEQYRISPKVMCGRMDDIP 364
DB      1 KDPDTPIRSPVMAAGGLFAVDRKMFWEIGGYDGLIWMGGEQYRISPKVMCGRMDDIP 60

QY      365 CSRVGHIYRKYVYKYPAGVS---LARNIKRVAEYVMDSYAEYIYQRRPEYHL--SAGD 419
DB      61 CSHVGHVFRKRSPTTPFGKSGSDVLSRNTVRVAEYVMDDYKRYFKHNPQAKRVDPGD 120

QY      420 VAVQKIRSSLNCKSPKFWMTKIAMDI 446
DB      121 ISERKEIRKELQCKSPKFWMTLVNYPDL 147

RESULT 31
: US-09-270-767-58910
: Sequence 58910, Application US/09270767
: Patent No. 6703491
: GENERAL INFORMATION:
: APPLICANT: Homburger et al.
: TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
: FILE REFERENCE: File Reference: 7326-094
: CURRENT APPLICATION NUMBER: US/09/270,767
: PRIOR FILING DATE: 1999-03-17
: NUMBER OF SEQ ID NOS: 62517
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 58910
: LENGTH: 161
: TYPE: PRT
: ORGANISM: Drosophila melanogaster
: US-09-270-767-58910
```

Query Match	8.5%;	Score 277.5;	DB 4;	Length 148;
Best Local Similarity	37.5%;	Pred. No. 1.9e-21;		

[illegible]

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 43809
LENGTH: 300
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-43809

Query Match
Best Local Similarity 30.0%; Pred. No. 2.6e-16;
Matches 63; Conservative 33; Mismatches 97; Indels 17; Gaps 6;

383 GVSANLAKRVAWMDV-VAEYTORPEYRHLSAGDVAVOKKLSLNCSPKPMFK 441
32 GPITTYNKKVITFPDTHKEFYTEPLARLYDNGDISQALAKRLKSKSFQWMDH 91
442 IAMDLPKFPVPPPAAMGEIRNVGTGLCADTKHGLSGPLRLGCGVRGGAAMNMQ 501
92 IADVDKFPGL-PANLHMGELRSVASDGLDSMGHPAIMGITVCHG-----GNNQ 144
502 VFTFTRREDIRPDGPHQTKKFCFDALSHSPVLYDCHSMKNQWKYRK-DKTLHPVS 560
145 LVRLNAGQLGVEGR-----CVEA--DROGIKLAVCRGLGVDPGPMQYNEHTGILMRVH 196
561 GSGMDCSESDHRIFMNTCNPSLSLQWLFEE 590
197 KCKMALHPATQQLSLGCHDVNDSDYQWMPK 226

RESULT 36

US-09-270-767-61773
Sequence 61773; Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 61773
LENGTH: 290
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-61773

Query Match
Best Local Similarity 7.0%; Score 228.5; DB 4; Length 290;
Matches 82; Conservative 31; Mismatches 95; Indels 79; Gaps 14;

345 EOYEISFKVMMCGRMEDIPCSRVGHIYRKVPYKVPAGVS--LARNLKRAVAMDEYA 402
1 ENWEMSRRIQCGGRVVISPCSHGHVFRSSTPTFFCGMSEVLTDLNLAATAVWMDW- 59
403 EYIYQRRPEYRHLSAGD---VAVOKKURSSLNCKSFWMFKIAMDLPKTFPYVEPPAA 459
60 QYFIMLYTSGITLGAQKDKVNTTERVALRBLQCKPFSWYLENI-W--PEHFFPA--PDRF 114
460 WGRIRNV-GTGLCA-----DTKHGALGSLRLAE--GCVR 490
115 FGKIILWDETECAQAYSKIMKNI-PGRALSRWKRAFEEDSKAEELMALIDLRDKCLR 174
491 -----GRGEAAMNMQVFTFTRREDIRPDG-----QHTK---KFCFD 525
175 PLKEDVPRSSLSAVTVGDSHAGSMOMFVITPRGQIMTNDVCLTYRQOKLGVYIKLKN 234
526 AISHTSVYLYDCHSMKNQWKYRKDKTLHPVYSGSCMDCSESDHR 572
235 RNATTSNVMLAOCAS--DSSQLMTYD-----MDTQOQISHR 267

RESULT 37

US-09-270-767-61989
Sequence 61989; Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 61989
LENGTH: 72
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-61989

Query Match
Best Local Similarity 6.7%; Score 218.5; DB 4; Length 72;
Matches 41; Conservative 11; Mismatches 20; Indels 1; Gaps 1;

218 IRTMLGASVATGDTTFLDSHCEANVNLPPILDIRANRKTIVCPMIDVIDHDDFRYE 277
1 IRARLGAHVSGEVITFLDACECTEGWLEPLARIVQNRRTVCPIIDVIDSETEFEYI 60
278 TQAGDAMRGAPFW 290
61 T-ASDSTWGFNW 72

RESULT 38

US-09-270-767-33350
Sequence 33350; Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 33350
LENGTH: 73
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33350

Query Match
Best Local Similarity 5.1%; Score 168; DB 4; Length 73;
Matches 29; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

352 KYMMCGRMEDIPCSRVGHIYRKVPYKVPAGVS--LARNLKRAVAMDEYAEYIYQR 409
12 QVMCGVLEIAPCSKVGHVFRKSTPTTFFGTTETIVNNARLVEVWMDWKEFYTSFY 71
410 P 410
72 P 72

RESULT 39

US-09-270-767-48567
Sequence 48567; Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
NUMBER OF SEQ ID NOS: 62517

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48567
LENGTH: 73
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-48567

Query Match 5.1%; Score 168; DB 4; Length 73;
Best Local Similarity 47.5%; Pred. No. 3.6e-10;
Matches 29; Conservative 11; Mismatches 19; Indels 2; Gaps 1;

QY 352 KVMCGRMEDICSRVGHVIRKYPKVPAGVS--LAARLKVAEVMDEVAEYIY 409
DB 12 QVMCGVLEIAPCSRGHVFRKSTYTFPGGTTEIYNHNNALVEWILDMKEFYSPY 71

QY 410 P 410
DB 72 P 72

RESULT 40
US-09-270-767-33441
Sequence 33441, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:

APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 33441
LENGTH: 101
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33441

Query Match 5.1%; Score 167; DB 4; Length 101;
Best Local Similarity 48.3%; Pred. No. 7.9e-10;
Matches 29; Conservative 12; Mismatches 17; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDICSRVGHVIRKYPKVPAGVS--LAARLKVAEVMDEVAEYIY 406
DB 26 INFQIWCGGILEITICSHVGHVFRDKSPYTFPGVAKIVLHNAARVAEVMDEWDFY 85

RESULT 41
US-09-270-767-48658
Sequence 48658, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48658
LENGTH: 101
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-48658

Query Match 5.1%; Score 167; DB 4; Length 101;
Best Local Similarity 48.3%; Pred. No. 7.9e-10;
Matches 29; Conservative 12; Mismatches 17; Indels 2; Gaps 1;

QY 349 ISFKVMCGRMEDICSRVGHVIRKYPKVPAGVS--LAARLKVAEVMDEVAEYIY 406
DB 26 INFQIWCGGILEITICSHVGHVFRDKSPYTFPGVAKIVLHNAARVAEVMDEWDFY 85

RESULT 42
US-09-134-000C-5947
Sequence 5947, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: Patentin version 3.1
SEQ ID NO 5947
LENGTH: 424
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-5947

Query Match 4.1%; Score 133; DB 4; Length 424;
Best Local Similarity 24.2%; Pred. No. 4.1e-05;
Matches 59; Conservative 35; Mismatches 96; Indels 54; Gaps 7;

QY 142 ETLPTNTSIIIPPHNGWSLRLTVSHVLSNRSPELVAEVLVDPSDREHL-----KK 194
DB 51 EDAPFVSIVPAHNEG-IYIVTVEALINFNYPQRIEIVINDNSDSASALLANVQCK 109

QY 195 PLEDYVALFPVSRIILTKRKREGILTRMKGASVATGDTVITFDISHCEANVMPLPLDRI 254
DB 110 AVNRMLQIINTDNITNGKGSNALN---IGFKQAKGDVIAIYDADNTPERNALRYLVGL 166

QY 255 -----ANRRTIYCPMIDVIDHDHDFRFEYQAGAMGAFPMWYKRIRIP 300
DB 167 IASBEYGAIVGKFRTRNRASLLTRFINETLAFQMGAGRF-----QLFKLCTIP 218

QY 301 PELQKADPEDPESPVMAGLPAVDKRMFELGVDYDGLIEWGGQYBISFKVMCGRM 360
DB 219 -----GTNFIYRSIIEALIGCHDDKRL---AEDTEISFRIYMGYKI 257

QY 361 EDIP 364
DB 258 KFP 261

RESULT 43
US-09-270-767-33601
Sequence 33601, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 33601
LENGTH: 88
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-33601

Query Match 3.9%; Score 129; DB 4; Length 88;
Best Local Similarity 34.7%; Pred. No. 7.8e-06;
Matches 26; Conservative 13; Mismatches 34; Indels 2; Gaps 2;

QY 531 SPVTLYDCHSGMGNOLWKY-RKDKTLHPVSG-SCMDCSESDHRIFMNTCNSSLTQWL 568
DB 9 APVWLDCHSGGNGQYWIYDYRHKLKGTGBRCLELLPFSQEVVANKCDYDNRFOQW 68

QY 589 FEHNTSTVLEKFNRN 603
| | | | |
| | | | |
Db 69 FGSFNKTALDNYSD 83

RESULT 44
US-09-270-767-48818
; Sequence 48818, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48818
; LENGTH: 88
; TYPE: PRT
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-48818

Query Match 3.9%; Score 129; DB 4; Length 88;
Best Local Similarity 34.7%; Pred. No. 7,8e-06;
Matches 26; Conservative 13; Mismatches 34; Indels 2; Gaps 2;

QY 531 SPVTLYCHSMKGNQMLKY-RKDKTYHPVSG-SCMDCSDHRIFNATCNPSSLTQML 588
| | | | |
| | | | |
Db 9 APVLMWCHSGGQVYDYRHKQLHGTGRCLBLPFSQEVVANKCDTDNRFOQWN 68
| | | | |
| | | | |
QY 589 FEHNTSTVLEKFNRN 603
| | | | |
| | | | |
Db 69 FGSFNKTALDNYSD 83

RESULT 45
US-09-973-457-5
; Sequence 5, Application US/09973457
; Patent No. 6703230
; GENERAL INFORMATION:
; APPLICANT: Kapellier-Libermann, Rosana
; TITLE OF INVENTION: 47174, A NOVEL HUMAN GLYCOSYLTRANSFERASE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 10448-099001
; CURRENT APPLICATION NUMBER: US/09/973,457
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/238,849
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-09-973-457-5

Query Match 3.9%; Score 128; DB 4; Length 135;
Best Local Similarity 30.4%; Pred. No. 2,1e-05;
Matches 42; Conservative 21; Mismatches 47; Indels 28; Gaps 10;

QY 468 TGLCAD---TKGALDSPLECGVRGGEAANNNOVFTFWREDIRPGDPQHTKKFC 523
| | | | |
| | | | |
Db 11 TGLCLDVNGSESKSDNPVQJLMDG-HGGGNQMLWK-----LTVNES--DGAIRINSDLC 61
| | | | |
| | | | |
QY 524 FDIASHSPVTLYCHSM-KG--NQLWKYRKDKTYHPV-----SGSCMDCSDHRI 573
| | | | |
| | | | |
Db 62 ---LTVNGTYLVISCDGTGKNDNQKWEVNNKDGITIRPKSKKGVDSGLCLDVKG--NKV 117
| | | | |
| | | | |
QY 574 FNNTCNPSL--TQOMLFE 590
| | | | |
| | | | |

Db 118 QLWTCNGSDAPNOKWIFE 135

RESULT 46
US-09-328-352-4416
; Sequence 4416, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4416
; LENGTH: 436
; TYPE: PRT
; ORGANISM: *Acinetobacter baumannii*
US-09-328-352-4416

Query Match 3.9%; Score 127; DB 4; Length 436;
Best Local Similarity 18.8%; Pred. No. 0.00019;
Matches 55; Conservative 60; Mismatches 102; Indels 76; Gaps 10;

QY 142 ETLPNTSIIIPFNEGMSLRTVHVSIVNRSPELVAEIYVDFSGRHLKKPLEDYMA 201
| | | | |
| | | | |
Db 67 ERMPVLSVLIIPAVNEG-VIEDTLHAIQODYAESEVLLINDGS-KDNTLEIAERLAK 124
| | | | |
| | | | |
QY 202 LFPSVILRTKK--REGILRTMLGASVATGVITFLDSHCANVWMLPPLDRIARNR 258
| | | | |
| | | | |
Db 125 IYPCIKIVNVPKMGKSKRTLNGLPHAKGSLIVYVADSTPEPDCVRLAQTLADK 184
| | | | |
| | | | |
QY 259 KTI-----VCMIDVYIHDPRYETQGDAMKGFDEMYYKRIPIPELO 304
| | | | |
| | | | |
Db 185 KLVAVNGKVRTWQDSILTRFIAIEIFPQWIFQGG-----RQOR----- 225
| | | | |
| | | | |
QY 305 KADSPDFESPVAAGLPAVDRKMPWELGGYDPLGIMWGEQYKISPKVMCGRMEDIP 364
| | | | |
| | | | |
Db 226 -----FELSTLMGNTVYIWRDALETGDFDEKSLV--DDTEMSFRITIGQKRIKWP 275
| | | | |
| | | | |
QY 365 CSRVG-----HTRKYVPYK-----PAGVSLARL 390
| | | | |
| | | | |
Db 276 YV-IGWQDPPLSLSVFVKQSRWTOGNFVYTRKLYPALRTPPIGIELNNI 327
| | | | |
| | | | |

RESULT 47
US-09-437-277-3
; Sequence 3, Application US/09437277
; Patent No. 6444447
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
; TITLE OF INVENTION: POLYMER GRAFTING BY POLYSACCHARIDE SYNTHASES
; FILE REFERENCE: 5820.551
; CURRENT APPLICATION NUMBER: US/09/437,277
; CURRENT FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: WordPerfect 8.0 (saved in ASCII format)
; SEQ ID NO 3
; LENGTH: 965
; TYPE: PRT
; ORGANISM: *Paeteurella multocida*
US-09-437-277-3

Query Match 3.8%; Score 125.5; DB 4; Length 965;
Best Local Similarity 20.9%; Pred. No. 0.001;
Matches 73; Conservative 48; Mismatches 133; Indels 95; Gaps 16;

QY 127 LPDIRHPN-----CNSKRYIETLP-----NTSIIIPFNEGMSLRTVHVSIVNRSPEL 176
| | | | |
| | | | |
Db 125 LPD-HVNDPTWYKRNKRSIGIRPVKNIGLSIIIPFNK--SRIDITLACVNOXTNY 180
| | | | |
| | | | |
QY 177 VAEIVLVDDPSDRHLKKPLEDYMALFPVSRIARTK-REGILRTMLGASVATGVITP 235
| | | | |
| | | | |

```

Db      181 PEEVVVADDS-KENLITVQKEOKL-DIKYVRQKDYQOLCAVNLGRTAKYFVSI 238
        236 LDBHCANVWMLPPLDRIARNKTIIVPMIDVID-----HDDPEYEQADAR 285
        239 LDDCMAPOOLWMSYLTLEDNDIVLIGPKRYVDHNTIAEPLNDPILIESLPETATN 298
Qy      286 G-----AFDWMYKRIPPELOKADSPDPESP--VMAGGLPAVDKRWELG 333
        299 NNPSTISKNIISLDWRLH-----FKTDLRLCDSPFRYFVAGNVAFSKEMLNKVG 350
Db      334 GYDPGLIENGGEQYEISFKWM--CGGRMED-----IP 364
        351 WPEEBFNHWGGEVEGEYRLFAGGCFRVIDGMAIHOBPGKENETEREAGKSTIKIV 410
Qy      365 CSRVGHIYRYVP-----YKVPAGVSLANLKRVAEVMWDEY--AEVI 405
        411 KEKVPYIRKLPIEDSHIRIP-----LVSITPAYNCANYI 448

```

```

RESULT 48
US-09-437-277-1
; Sequence 1, Application US/09437277
; Patent No. 6444447
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF REGENTS OF THE UNIVERSITY OF OKLAHOMA
; TITLE OF INVENTION: POLYMER GRAFTING BY POLYSACCHARIDE SYNTHASES
; FILE REFERENCE: 5820.551
; CURRENT APPLICATION NUMBER: US/09/437, 277
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: WordPerfect 8.0 (saved in ASCII format)
; LENGTH: 702
; TYPE: PRF
; ORGANISM: Pasteurella multocida
US-09-437-277-1

```

```

Query Match      3.8%; Score 124; DB 4; Length 702;
Best Local Similarity 22.1%; Pred. No. 0.0089;
Matches 83; Conservative 49; Mismatches 156; Indels 88; Gaps 19;

Qy      61 KTFELDGOK--LKD-WH--DKEAIRDAQVNGEGRPMTDAERVDQVRENGFN 114
        84 KKLVLSDSEKNTLKNWKLLTEKSENAEVAVALVXPDPKDLVAPLDHV---NDF 140
Db      115 IYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGWSLLRVHSLRSP 174
        141 WYKRRKRLGIR-PEHQ-----VGLSITVTFR--PALSTTLACLNVQKT 185
Qy      175 ELVAEIVLVDDFSDREHLKKPLDYMALPSPVRLRTKKREG--LRTKMLGASVATGV 232
        186 HYFEIYIVDDG--QEDLSIIRQYENKL-DIRYVR-QKDNQQAARMGRLAKYDF 242
Db      233 IFFLDHSCANVWMLPPLDRIARNK-TIVCP--MIDV--IDHDF-----RY 276
        243 IGLDDCMAPNPLWMSYVALEDDDLITIGPKYIDTQHIDPKDFLNNASSLLESLEY 302
Qy      277 ETQAGDAMG---ADWMYKRIPIPELOKADSPDPESP--VMAGGLPAVDKRWEL 330
        303 KTNNSVAAGGEGTSLDMWL-----EQFEKTENLRSDSPFRFAGNVAFARKWLN 354
Db      331 ELGGYDPGLIENGGEQYEISFKVMCGGRME-----361
        355 KSGPFBEBFNHWGGEVEGEYRLFYGSFKTIDGIMAHOBPGKENETEREAGKNTIL 414
Qy      362 DIPCSRVGHIYRYVP 377
        415 DIMREKVPYIRKLIP 430

```

RESULT 49
US-09-134-000C-5087

```

; Sequence 5087, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucelte-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5087
; LENGTH: 534
; TYPE: PRF
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5087

```

```

Query Match      3.8%; Score 123.5; DB 4; Length 534;
Best Local Similarity 22.7%; Pred. No. 0.0064;
Matches 56; Conservative 43; Mismatches 93; Indels 55; Gaps 11;

Qy      145 PNTSIIIPPHN--EGWSLLRTVHSLVNSPPELVAEIYLVDDFSDREHLKKPLDYMAL 202
        193 PXSIMAPYVNEBKWLRL--CIDSILNQVYTNW--ELCMADASTDPVVKILTEYQOL 248
Db      203 PPSVRLRTKKREGIIRTMGLASVATGVITFLDHCANVWMLPPLDRIARNKTIIV 262
        249 DERIRVFERQNGHISEATNSALATGFEVALLDDELAINAYEVVKLVNEN----- 303
Qy      263 CEMIDVIDHDPRYETQADAMRGAFDWMYKRIPIPELOKADPS--DPESPVMAGG- 320
        304 -PELDLIVSDEDKID-----NDG-----NRSDPAFRDMSPLDLGT 339
Db      321 ----LFAVDRKFMWELGYPGLEIENGGEQYEISFKVMCGGRMEDIPCSVGH---- 371
        340 NYISHLGVYRRSLIERIGFRKGYE--GSQDYDLVLRF-----TEKTKERIKIIPKVL 391
Qy      372 -YRKVYP 377
        392 YTRWLP 398

```

```

RESULT 50
US-09-107-532A-6889
; Sequence 6889, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucelte-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997

```


Inis Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 14:04:27 ; Search time 147 Seconds

(without alignments)
1467.854 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278
Sequence: 1 MRKEKRLQAVLVLAALV.....TQOWLEHTNSTVLEKENN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1589859 seqs, 357834939 residues

Total number of hits satisfying chosen parameters: 1589859

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubppa/PCF_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubppa/PCFUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubppa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubppa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubppa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubppa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubppa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/1/pubppa/US11_NEW_PUB.pep:*
- 19: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3278	100.0	603	US-09-795-926-43	Sequence 43, Appl
2	3278	100.0	603	US-10-001-851-2	Sequence 2, Appl
3	3278	100.0	603	US-10-364-774-43	Sequence 43, Appl
4	3242.5	98.9	631	US-09-795-926-41	Sequence 41, Appl
5	3242.5	98.9	631	US-10-364-774-41	Sequence 41, Appl
6	2771	84.5	506	US-09-795-926-31	Sequence 31, Appl
7	2771	84.5	506	US-10-364-774-31	Sequence 31, Appl
8	2746.5	83.8	535	US-09-795-926-29	Sequence 29, Appl
9	2746.5	83.8	535	US-10-364-774-29	Sequence 29, Appl
10	1877	57.3	366	US-09-795-926-39	Sequence 39, Appl
11	1877	57.3	366	US-10-364-774-39	Sequence 39, Appl
12	1392	42.5	276	US-10-292-896-3	Sequence 3, Appl
13	1370	41.8	269	US-09-795-926-27	Sequence 27, Appl

14	1370	41.8	269	US-10-364-774-27	Sequence 27, Appl
15	1311	40.0	321	US-09-795-926-35	Sequence 35, Appl
16	1311	40.0	321	US-10-364-774-35	Sequence 35, Appl
17	1125	34.3	561	US-09-925-301-1006	Sequence 1006, Ap
18	1117	34.1	559	US-10-001-851-24	Sequence 24, Appl
19	1116	34.0	559	US-10-001-851-21	Sequence 21, Appl
20	1116	34.0	559	US-10-205-219-36	Sequence 36, Appl
21	1116	34.0	559	US-10-205-219-76	Sequence 76, Appl
22	1115	34.0	559	US-10-001-851-22	Sequence 22, Appl
23	1115	34.0	559	US-10-001-851-23	Sequence 23, Appl
24	1113	34.0	559	US-10-001-851-20	Sequence 20, Appl
25	1106	33.7	556	US-10-292-896-62	Sequence 62, Appl
26	1079	32.9	626	US-10-001-851-27	Sequence 27, Appl
27	1050	32.0	187	US-10-292-896-118	Sequence 118, App
28	1033	31.5	578	US-10-074-527-8	Sequence 8, Appl
29	1029	31.4	657	US-10-341-434-32	Sequence 32, Appl
30	1013	30.9	940	US-10-821-273-86	Sequence 86, Appl
31	999.5	30.5	558	US-10-292-896-66	Sequence 66, Appl
32	999.5	30.5	558	US-10-433-256-6	Sequence 6, Appl
33	991.5	30.2	581	US-10-074-527-2	Sequence 2, Appl
34	991.5	30.2	581	US-10-085-198-12	Sequence 122, App
35	991.5	30.2	581	US-10-433-256-4	Sequence 4, Appl
36	990	30.2	581	US-10-789-241-44	Sequence 44, Appl
37	990	30.2	240	US-09-795-926-33	Sequence 33, Appl
38	990	30.2	240	US-10-364-774-33	Sequence 33, Appl
39	988.5	30.2	612	US-10-001-851-25	Sequence 25, Appl
40	988	30.1	209	US-09-795-926-37	Sequence 37, Appl
41	988	30.1	209	US-10-364-774-37	Sequence 37, Appl
42	971.5	29.6	581	US-10-292-896-58	Sequence 58, Appl
43	971	29.6	552	US-09-815-028-2	Sequence 2, Appl
44	971	29.6	552	US-10-028-072-196	Sequence 196, App
45	971	29.6	552	US-10-140-808-196	Sequence 196, App

ALIGNMENTS

RESULT 1
US-09-795-926-43
Sequence 43, Application US/09795926
Patent No. US2002098486A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Brian
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wilgenowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kieker, James Alvin
APPLICANT: Poter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 603
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-43

Query Match 100.0%; Score 3278; DB 9; Length 603;
Best Local Similarity 100.0%; Pred. No. 5e-310;
Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODPTGGSGAAYAPAAAGGSHSRK 60
DB 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODPTGGSGAAYAPAAAGGSHSRK 60
QY 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
DB 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
QY 121 ISLNSRLPIRHPNCNKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRPPELVABI 180
DB 121 ISLNSRLPIRHPNCNKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRPPELVABI 180
QY 181 VLVDPDSDBEHLKKPLEDMALPPSVRIIRTKRBSLIRTRMLGASVATGDIVTFLDSDC 240
DB 181 VLVDPDSDBEHLKKPLEDMALPPSVRIIRTKRBSLIRTRMLGASVATGDIVTFLDSDC 240
QY 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMEMYKRIPIR 300
DB 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMEMYKRIPIR 300
QY 301 PELQKADPSDPFSPVMAAGLPAVDRKMFELGCGYDPCGLEIWGEGYEISFKVMCGGM 360
DB 301 PELQKADPSDPFSPVMAAGLPAVDRKMFELGCGYDPCGLEIWGEGYEISFKVMCGGM 360
QY 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRAEVMDEYAEYIYQRPBYRHLSAGDV 420
DB 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRAEVMDEYAEYIYQRPBYRHLSAGDV 420
QY 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGRIANVTGLCADTKGALG 480
DB 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGRIANVTGLCADTKGALG 480
QY 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRGDPQHTKKCFDAISHTSPVTLVDCS 540
DB 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRGDPQHTKKCFDAISHTSPVTLVDCS 540
QY 541 MGNOLMKYRKDXTLYHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSTVLEKF 600
DB 541 MGNOLMKYRKDXTLYHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 2
US-10-001-851-2
/ Sequence 2, Application US/10001851
/ Publication No. US20020115628A1
/ GENERAL INFORMATION:
/ APPLICANT: MEYERS, Rachel A.
/ APPLICANT: WILLIAMSON, Mark
/ TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1el Human Glycosyl Transferase
/ TITLE OF INVENTION: Uses Thereof
/ FILE REFERENCE: 10147-56U1
/ CURRENT APPLICATION NUMBER: US/10/001,851
/ PRIOR FILING DATE: 2001-11-20
/ PRIOR APPLICATION NUMBER: US 60/249,939
/ PRIOR FILING DATE: 2000-11-20
/ NUMBER OF SEQ ID NOS: 29
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 603
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-001-851-2

Query Match 100.0%; Score 3278; DB 13; Length 603;
Best Local Similarity 100.0%; Pred. No. 5e-310;

Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODPTGGSGAAYAPAAAGGSHSRK 60
DB 1 MRKREKRLLOAVLVAALVLLPNVGLMALYRERODPTGGSGAAYAPAAAGGSHSRK 60
QY 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
DB 61 KTFPLDGGKLDKMDHKEAIRDAQRVNGEGRPYPMTDARVDAQYRENGFNIVSDK 120
QY 121 ISLNSRLPIRHPNCNKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRPPELVABI 180
DB 121 ISLNSRLPIRHPNCNKRYLETLPNTSIIIPHNAGWSLLRTVSHVNSRPPELVABI 180
QY 181 VLVDPDSDBEHLKKPLEDMALPPSVRIIRTKRBSLIRTRMLGASVATGDIVTFLDSDC 240
DB 181 VLVDPDSDBEHLKKPLEDMALPPSVRIIRTKRBSLIRTRMLGASVATGDIVTFLDSDC 240
QY 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMEMYKRIPIR 300
DB 241 EAVNVMPLPLDLRIANRKTIVCPMIDVIDHDDFRYETQAGDAMRGAPDMEMYKRIPIR 300
QY 301 PELQKADPSDPFSPVMAAGLPAVDRKMFELGCGYDPCGLEIWGEGYEISFKVMCGGM 360
DB 301 PELQKADPSDPFSPVMAAGLPAVDRKMFELGCGYDPCGLEIWGEGYEISFKVMCGGM 360
QY 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRAEVMDEYAEYIYQRPBYRHLSAGDV 420
DB 361 EDIPCSRVGHIYRKVPYVPAGVSLARNLKRAEVMDEYAEYIYQRPBYRHLSAGDV 420
QY 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGRIANVTGLCADTKGALG 480
DB 421 AVQKRLSSLNCKSFPMFTKIAMDLPKFYPPVEPAAAMGRIANVTGLCADTKGALG 480
QY 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRGDPQHTKKCFDAISHTSPVTLVDCS 540
DB 481 SPLRLGECYRGGEAAMNNMQUFTFTWRREDIRGDPQHTKKCFDAISHTSPVTLVDCS 540
QY 541 MGNOLMKYRKDXTLYHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSTVLEKF 600
DB 541 MGNOLMKYRKDXTLYHPVSGSCMDCSESDHRIFMNTCNPSLTOOMLFEHTNSTVLEKF 600
QY 601 NRN 603
DB 601 NRN 603

RESULT 3
US-10-364-774-43
/ Sequence 43, Application US/10364774
/ Publication No. US20030144497A1
/ GENERAL INFORMATION:
/ APPLICANT: Donoho, Gregory
/ APPLICANT: Hilbun, Erin
/ APPLICANT: Turner, C. Alexander Jr.
/ APPLICANT: Friedrich, Glenn
/ APPLICANT: Abuin, Alejandro
/ APPLICANT: Zambrowicz, Brian
/ APPLICANT: Sande, Arthur T.
/ APPLICANT: Walke, D. Wade
/ APPLICANT: Wilganowski, Nathaniel L.
/ APPLICANT: Hu, Yi
/ APPLICANT: Kleke, James Alvin
/ APPLICANT: Potter, David George
/ TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
/ TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
/ FILE REFERENCE: LEX-0144-USA
/ CURRENT APPLICATION NUMBER: US/10/364,774
/ PRIOR FILING DATE: 2003-02-11
/ PRIOR APPLICATION NUMBER: US/09/795,926
/ PRIOR FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: US 60/185,920
/ PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 43
 ; LENGTH: 603
 ; TYPE: PRF
 ; ORGANISM: homo sapiens
 ; US-10-364-774-43

Query Match 100.0%; Score 3278; DB 14; Length 603;
 Best Local Similarity 100.0%; Pred. No. 5e-310;
 Matches 603; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRRERKRLIQAVALVLAALVLLPNVGLMALYERQDPTGGSGAAVAPAGGSHSRK 60
 DB 1 MRRERKRLIQAVALVLAALVLLPNVGLMALYERQDPTGGSGAAVAPAGGSHSRK 60
 QY 61 KTFPLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGFNIVYSDK 120
 DB 61 KTFPLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGFNIVYSDK 120
 QY 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNCGMSLLRTVSHVLSRSPPELVAEI 180
 DB 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNCGMSLLRTVSHVLSRSPPELVAEI 180
 QY 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 DB 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 QY 241 EAVVNMPLPLLDRIARNRKTIVCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 241 EAVVNMPLPLLDRIARNRKTIVCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 QY 301 PELQKADPSDPFSPVWAGGLFAVDRKFMELGYPGLIINGGEOYEISFKYMMCGGRM 360
 DB 301 PELQKADPSDPFSPVWAGGLFAVDRKFMELGYPGLIINGGEOYEISFKYMMCGGRM 360
 QY 361 EDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRREYHLSAGDV 420
 DB 361 EDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRREYHLSAGDV 420
 QY 421 AVQKULRSSLNCSEFKFMFTKIAMDLKFPYPPVBPAAANGERTINVTGICADTKGALG 480
 DB 421 AVQKULRSSLNCSEFKFMFTKIAMDLKFPYPPVBPAAANGERTINVTGICADTKGALG 480
 QY 481 SPLRLEGCVARGRAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 DB 481 SPLRLEGCVARGRAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 QY 541 MKNQULMKYRKDKTLYHPVSGSCMDCESDHRIFMNTCNPSLSLQOWLFEHTNSTVLEKF 600
 DB 541 MKNQULMKYRKDKTLYHPVSGSCMDCESDHRIFMNTCNPSLSLQOWLFEHTNSTVLEKF 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 4
US-09-795-926-41

; Sequence 41, Application US/09795926
 ; Patent No. US20020098486a1
 ; GENERAL INFORMATION:
 ; APPLICANT: Donoho, Gregory
 ; APPLICANT: Hilbun, Brian
 ; APPLICANT: Turner, C. Alexander Jr.
 ; APPLICANT: Friedrich, Glenn
 ; APPLICANT: Abudin, Alejandro
 ; APPLICANT: Zambrowicz, Brian
 ; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade
 ; APPLICANT: Wilgenowski, Nathaniel L.
 ; APPLICANT: Hu, Yi
 ; APPLICANT: Kieke, James Alvin
 ; APPLICANT: Potter, David George
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
 ; FILE REFERENCE: LEX-0144-USA
 ; CURRENT APPLICATION NUMBER: US/09/795,926
 ; CURRENT FILING DATE: 2001-02-28
 ; PRIOR APPLICATION NUMBER: US 60/185,920
 ; PRIOR FILING DATE: 2000-02-29
 ; PRIOR APPLICATION NUMBER: US 60/186,558
 ; PRIOR FILING DATE: 2000-03-02
 ; PRIOR APPLICATION NUMBER: US 60/191,849
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 47
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 41
 ; LENGTH: 631
 ; TYPE: PRF
 ; ORGANISM: homo sapiens
 ; US-09-795-926-41

Query Match 98.9%; Score 3242.5; DB 9; Length 631;
 Best Local Similarity 95.4%; Pred. No. 1.6e-306;
 Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 1 MRRERKRLIQAVALVLAALVLLPNVGLMALYERQDPTGGSGAAVAPAGGSHSRK 60
 DB 1 MRRERKRLIQAVALVLAALVLLPNVGLMALYERQDPTGGSGAAVAPAGGSHSRK 60
 QY 61 KTFPLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGFNIVYSDK 120
 DB 61 KTFPLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDARVDAQYRENGFNIVYSDK 120
 QY 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNCGMSLLRTVSHVLSRSPPELVAEI 180
 DB 121 ISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNCGMSLLRTVSHVLSRSPPELVAEI 180
 QY 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 DB 181 VLVDPSDBREHLKKPLEDYMALFPVSRIILRTKKREGILRTMLGASVATGDTITFLDSHC 240
 QY 241 EAVVNMPLPLLDRIARNRKTIVCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 DB 241 EAVVNMPLPLLDRIARNRKTIVCPMTDIVDHDFFRYETQAGDAMRGAFPMEMYKRIPI 300
 QY 301 PELQKADPSDPFSPVWAGGLFAVDRKFMELGYPGLIINGGEOYEISFKYMMCGGRM 360
 DB 301 PELQKADPSDPFSPVWAGGLFAVDRKFMELGYPGLIINGGEOYEISFKYMMCGGRM 360
 QY 361 VSNMSPQAVFLPRAAPMLALQVMMCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNL 420
 DB 361 VSNMSPQAVFLPRAAPMLALQVMMCGGRMEDIPCSRVGHIYRKYVPYKVPAGVSLARNL 420
 QY 421 RVAEVMDEYAEIYQRRREYHLSAGDVAVQKULRSSLNCSEFKFMFTKIAMDLKFPY 480
 DB 421 RVAEVMDEYAEIYQRRREYHLSAGDVAVQKULRSSLNCSEFKFMFTKIAMDLKFPY 480
 QY 481 SPLRLEGCVARGRAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 DB 481 SPLRLEGCVARGRAANNNOVFTWRREDIRPGDPQHTKKPCPDASHTSPTLYDCHS 540
 QY 541 RPDGPQHTKKPCPDASHTSPTLYDCHSMKQULMKYRKDKTLYHPVSGSCMDCESDH 571
 DB 541 RPDGPQHTKKPCPDASHTSPTLYDCHSMKQULMKYRKDKTLYHPVSGSCMDCESDH 571
 QY 572 RIFMNTCNPSLSLQOWLFEHTNSTVLEKEN 601
 DB 601 RIFMNTCNPSLSLQOWLFEHTNSTVLEKEN 601

RESULT 5

US-10-364-774-41
; Sequence 41, Application US/10364774
; Publication No. US2003014497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/10/364,774
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 631
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-364-774-41

Query Match

Best Local Similarity 98.9%; Score 3242.5; DB 14; Length 631;
Matches 601; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 1 MRREKRLQVALVLAALVLPVNGIMALYRERODPTGGSGAAYAPAGGSHSRQK 60
DB 1 MRREKRLQVALVLAALVLPVNGIMALYRERODPTGGSGAAYAPAGGSHSRQK 60
QY 61 KTFELDGGQKLKMDHDEAIRDAQRYNGEGRPYMTAEVDDAYRENGNIYVSDK 120
DB 61 KTFELDGGQKLKMDHDEAIRDAQRYNGEGRPYMTAEVDDAYRENGNIYVSDK 120
QY 121 ISLNRSLPDIRHBNCSKRYLETLPNTSIIIPHNESMSILRTVSHVLSRSPPELVAE 180
DB 121 ISLNRSLPDIRHBNCSKRYLETLPNTSIIIPHNESMSILRTVSHVLSRSPPELVAE 180
QY 181 VLVDPSDREHLKKPLEDYMALFPSVRIILTKKREGIIRTRMLGASVATGDTITFLDSHC 240
DB 181 VLVDPSDREHLKKPLEDYMALFPSVRIILTKKREGIIRTRMLGASVATGDTITFLDSHC 240
QY 241 EAVVNMPLPLLDIAIRNRKTIYVCPMIDVIDHDPFRYETQGDAMRGAFFMEYTKKIP 300
DB 241 EAVVNMPLPLLDIAIRNRKTIYVCPMIDVIDHDPFRYETQGDAMRGAFFMEYTKKIP 300
QY 301 PELQKADPSDFSPVWAGLFAVDRKMFELGSGYDPGLEIMWGEQYEISFKLHMLPRL 360
DB 301 PELQKADPSDFSPVWAGLFAVDRKMFELGSGYDPGLEIMWGEQYEISFKLHMLPRL 360
QY 353 -----VMMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLK 391
DB 353 -----VMMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLK 391
QY 392 RAEVNMDEYAEIYQRRPEYRHLSDGDAVAVOKKLSSLNCKSPKFMFTKIANDLRFYF 451
DB 392 RAEVNMDEYAEIYQRRPEYRHLSDGDAVAVOKKLSSLNCKSPKFMFTKIANDLRFYF 451
QY 421 RAEVNMDEYAEIYQRRPEYRHLSDGDAVAVOKKLSSLNCKSPKFMFTKIANDLRFYF 480
DB 421 RAEVNMDEYAEIYQRRPEYRHLSDGDAVAVOKKLSSLNCKSPKFMFTKIANDLRFYF 480

RESULT 6

US-09-795-926-31
; Sequence 31, Application US/09795926
; Patent No. US20020098486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedlich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 506
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-31

Query Match

Best Local Similarity 84.5%; Score 2771; DB 9; Length 506;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDDAYRENGNIYVSDKISLNRSLPDIRHBNCSKRYLETLPNTSIIIPHNES 157
DB 1 MTDARVDDAYRENGNIYVSDKISLNRSLPDIRHBNCSKRYLETLPNTSIIIPHNES 157
QY 158 WSSLRTVSHVLSRSPPELVAEIVLVDPSDREHLKKPLEDYMALFPSVRIILTKKREG 217
DB 158 WSSLRTVSHVLSRSPPELVAEIVLVDPSDREHLKKPLEDYMALFPSVRIILTKKREG 217
QY 218 IRTRMGLASVATGDTITFLDSHCANVNMPLPLLDIAIRNRKTIYVCPMIDVIDHDPFRY 277
DB 218 IRTRMGLASVATGDTITFLDSHCANVNMPLPLLDIAIRNRKTIYVCPMIDVIDHDPFRY 277
QY 278 TQAGDAMRGAFFMEYTKKIPPELQKADPSDFSPVWAGLFAVDRKMFELGSGYDP 337
DB 278 TQAGDAMRGAFFMEYTKKIPPELQKADPSDFSPVWAGLFAVDRKMFELGSGYDP 337
QY 338 GLEIWGEQYEISFKVMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLKRYAEVW 397
DB 338 GLEIWGEQYEISFKVMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLKRYAEVW 397
QY 421 GLEIWGEQYEISFKVMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLKRYAEVW 480
DB 421 GLEIWGEQYEISFKVMCGRMEDIPCSRVGHIYRKVYPYKVPAGVSLARNLKRYAEVW 480

QY 398 MDEAYIYORREPEYHLSAGDVAVOKLRSSLNCKSFKMFMTKIAMDLPKFYPPVEPPA 457
DB 301 MDEAYIYORREPEYHLSAGDVAVOKLRSSLNCKSFKMFMTKIAMDLPKFYPPVEPPA 360
QY 458 AAMGEIRNVGTGICADTKHGALGSPRLBEGCVRGGEAAMNNQVFTFWREDIRGDDPQ 517
DB 361 AAMGEIRNVGTGICADTKHGALGSPRLBEGCVRGGEAAMNNQVFTFWREDIRGDDPQ 420
QY 518 HTKKFCFDAISHTSPVTLVDCHSMKGNQLMKRYKDKTLVHPVSGSCMDSESHPRIFMNT 577
DB 421 HTKKFCFDAISHTSPVTLVDCHSMKGNQLMKRYKDKTLVHPVSGSCMDSESHPRIFMNT 480
QY 578 CNPSSLTQOQWLFHTNSTVLEKFNRN 603
DB 481 CNPSSLTQOQWLFHTNSTVLEKFNRN 506

RESULT 7

US-10-364-774-31
Sequence 31, Application US/10364774
Publication No. US2003014497A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abulin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Poter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/10/364,774
CURRENT FILING DATE: 2003-02-11
PRIOR APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 31
LENGTH: 506
TYPE: PRT
ORGANISM: homo sapiens
US-10-364-774-31

Query Match 84.5%; Score 2771; DB 14; Length 506;
Best Local Similarity 100.0%; Pred. No. 1,le-260;
Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157
DB 1 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60
QY 158 WSSLIRTVHSLNRSPELVAEIVLVDDFSDBHLLKKPLEDYALPPSVRIILRTKREG 217
DB 61 WSSLIRTVHSLNRSPELVAEIVLVDDFSDBHLLKKPLEDYALPPSVRIILRTKREG 120
QY 218 IRTMLGASVATGDTITFLDSHCSEANVMWLPPLDIRIARRKTIIVCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTITFLDSHCSEANVMWLPPLDIRIARRKTIIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFDMWYKRIPIPELOKADSPFESPVWAGGLFAVDRKMFWEIGYDP 337

DB 181 TQAGDMRGAFFDMWYKRIPIPELOKADSPDFESPVMAGGLFAVDRKMFWEIGYDP 240
QY 338 GLBIEWGEOYEISFKWMMCGRMEDIPCSRVGHIYRKYPYKVPAGVSLARRNLKRYAEW 397
DB 241 GLBIEWGEOYEISFKWMMCGRMEDIPCSRVGHIYRKYPYKVPAGVSLARRNLKRYAEW 300
QY 398 MDEAYIYORREPEYHLSAGDVAVOKLRSSLNCKSFKMFMTKIAMDLPKFYPPVEPPA 457
DB 301 MDEAYIYORREPEYHLSAGDVAVOKLRSSLNCKSFKMFMTKIAMDLPKFYPPVEPPA 360
QY 458 AAMGEIRNVGTGICADTKHGALGSPRLBEGCVRGGEAAMNNQVFTFWREDIRGDDPQ 517
DB 361 AAMGEIRNVGTGICADTKHGALGSPRLBEGCVRGGEAAMNNQVFTFWREDIRGDDPQ 420
QY 518 HTKKFCFDAISHTSPVTLVDCHSMKGNQLMKRYKDKTLVHPVSGSCMDSESHPRIFMNT 577
DB 421 HTKKFCFDAISHTSPVTLVDCHSMKGNQLMKRYKDKTLVHPVSGSCMDSESHPRIFMNT 480
QY 578 CNPSSLTQOQWLFHTNSTVLEKFNRN 603
DB 481 CNPSSLTQOQWLFHTNSTVLEKFNRN 506

RESULT 8

US-09-795-926-29
Sequence 29, Application US/09795926
Patent No. US20020098486A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abulin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Poter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 535
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-29

Query Match 83.8%; Score 2746.5; DB 9; Length 535;
Best Local Similarity 94.6%; Pred. No. 2,8e-258;
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;

QY 98 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 157
DB 1 MTAERVDQAYRENGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEG 60
QY 158 WSSLIRTVHSLNRSPELVAEIVLVDDFSDBHLLKKPLEDYALPPSVRIILRTKREG 217
DB 61 WSSLIRTVHSLNRSPELVAEIVLVDDFSDBHLLKKPLEDYALPPSVRIILRTKREG 120
QY 218 IRTMLGASVATGDTITFLDSHCSEANVMWLPPLDIRIARRKTIIVCPMIDVIDHDDFRYE 277

DB 121 IRTMIGASVATGDTVTFDLSHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 180
QY 278 TOAGDMRGAFDWMEMYKRIPIPELOKADPSDPFSPVWAGGLFVDRKFMWELGGYDP 337
DB 181 TOAGDMRGAFDWMEMYKRIPIPELOKADPSDPFSPVWAGGLFVDRKFMWELGGYDP 240
QY 338 GLEIWGGEQYEISFK-----VWCGRMEDIPCISRY 368
DB 241 GLEIWGGEQYEISFKGLHMLPRLVSNWPQAVFLPAPRMALQVWCGRMEDIPCISRY 300
QY 369 GHYRYKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDVAVOCKLRS 428
DB 301 GHYRYKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDVAVOCKLRS 360
QY 429 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGALGSLPLRLEG 488
DB 361 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGALGSLPLRLEG 420
QY 489 VGRGSAAMNNQVFTFTWRREDIRPGDPQHTKKFCFCDALSHTSPTLYDCHSMKGNQMLK 548
DB 421 VGRGSAAMNNQVFTFTWRREDIRPGDPQHTKKFCFCDALSHTSPTLYDCHSMKGNQMLK 480
QY 549 YRKDKTLNHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFEHTNSTVLEKENRN 603
DB 481 YRKDKTLNHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFEHTNSTVLEKENRN 535

RESULT 9

US-10-364-774-29
Sequence 29, Application US/10364774
Publication No. US2003014497A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/10/364,774
CURRENT FILING DATE: 2003-02-11
PRIOR APPLICATION NUMBER: US/09/795,926
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 535
TYPE: PRT
ORGANISM: homo sapiens
US-10-364-774-29

Query Match 83.8%; Score 2746.5; DB 14; Length 535;
Best Local Similarity 94.6%; Pred. No. 2.8e-258;
Matches 506; Conservative 0; Mismatches 0; Indels 29; Gaps 1;
QY 98 MDAERVDQAYRNGRIYVSDKISLNRSLPDIRHNGSKRYLETLPNTSIIPPHNAG 157
DB 1 MDAERVDQAYRNGRIYVSDKISLNRSLPDIRHNGSKRYLETLPNTSIIPPHNAG 60

QY 158 WSSLRTVHSVLNRSPELVAEIVLVDDPSDREHLKKPLEDMYALPPSVRIIRTKREGI 217
DB 61 WSSLRTVHSVLNRSPELVAEIVLVDDPSDREHLKKPLEDMYALPPSVRIIRTKREGI 120
QY 218 IRTMIGASVATGDTVTFDLSHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 277
DB 121 IRTMIGASVATGDTVTFDLSHCEANVMWLPLLDRIARNRTIYCPMIDVIDHDDFRYE 180
QY 278 TOAGDMRGAFDWMEMYKRIPIPELOKADPSDPFSPVWAGGLFVDRKFMWELGGYDP 337
DB 181 TOAGDMRGAFDWMEMYKRIPIPELOKADPSDPFSPVWAGGLFVDRKFMWELGGYDP 240
QY 338 GLEIWGGEQYEISFK-----VWCGRMEDIPCISRY 368
DB 241 GLEIWGGEQYEISFKGLHMLPRLVSNWPQAVFLPAPRMALQVWCGRMEDIPCISRY 300
QY 369 GHYRYKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDVAVOCKLRS 428
DB 301 GHYRYKYVPYKYPAGVSLARNLKRVAEVMWDEYAETIYORRPEYRHLASGDVAVOCKLRS 360
QY 429 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGALGSLPLRLEG 488
DB 361 SLNCKSFKFMFTKIAMDLPKFPYPPVPPAAWGEIRNVGTGLCADTKHGALGSLPLRLEG 420
QY 489 VGRGSAAMNNQVFTFTWRREDIRPGDPQHTKKFCFCDALSHTSPTLYDCHSMKGNQMLK 548
DB 421 VGRGSAAMNNQVFTFTWRREDIRPGDPQHTKKFCFCDALSHTSPTLYDCHSMKGNQMLK 480
QY 549 YRKDKTLNHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFEHTNSTVLEKENRN 603
DB 481 YRKDKTLNHPVSGSCMDCESDHRIFMNTCNPSLTOQWLFEHTNSTVLEKENRN 535

RESULT 10

US-09-795-926-39
Sequence 39, Application US/09795926
Patent No. US20020098486A1
GENERAL INFORMATION:
APPLICANT: Donoho, Gregory
APPLICANT: Hilbun, Erin
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Friedrich, Glenn
APPLICANT: Abuin, Alejandro
APPLICANT: Zambrowicz, Brian
APPLICANT: Sands, Arthur T.
APPLICANT: Walke, D. Wade
APPLICANT: Wiganowski, Nathaniel L.
APPLICANT: Hu, Yi
APPLICANT: Kleke, James Alvin
APPLICANT: Potter, David George
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
FILE REFERENCE: LEX-0144-USA
CURRENT APPLICATION NUMBER: US/09/795,926
CURRENT FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: US 60/185,920
PRIOR FILING DATE: 2000-02-29
PRIOR APPLICATION NUMBER: US 60/186,558
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 60/191,849
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 47
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 39
LENGTH: 366
TYPE: PRT
ORGANISM: homo sapiens
US-09-795-926-39

Query Match 57.3%; Score 1877; DB 9; Length 366;
Best Local Similarity 100.0%; Pred. No. 9.6e-174;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRERKRLQAVLVLAALVLLPVNGMALYREROPGTGGGGAAPAAAGGSHSRK 60
DB 1 MRERKRLQAVLVLAALVLLPVNGMALYREROPGTGGGGAAPAAAGGSHSRK 60
QY 61 KTFELDGQQLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDAQYRENGFNIVYSDK 120
DB 61 KTFELDGQQLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDAQYRENGFNIVYSDK 120
QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGWSSLLRTVSHVLRSPPELVAEI 180
DB 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGWSSLLRTVSHVLRSPPELVAEI 180
QY 181 VLVDSDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMIGASVATGDTVTFPLDSHC 240
DB 181 VLVDSDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMIGASVATGDTVTFPLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYYKRIPIP 300
DB 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYYKRIPIP 300
QY 301 PELQKADPSDPFESSPVMAAGLFPAVDRKFWELGSDYDGLIHWGEQYEISFKV 353
DB 301 PELQKADPSDPFESSPVMAAGLFPAVDRKFWELGSDYDGLIHWGEQYEISFKV 353

RESULT 11
US-10-364-774-39
; Sequence 39, Application US/10364774
; Publication No. US2003014497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuhin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wiganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kleke, James Alvin
; APPLICANT: Poter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/10/364, 774
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US/09/795, 926
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185, 920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186, 558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191, 849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 366
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-364-774-39

Query Match 57.3%; Score 1877; DB 14; Length 366;
Best Local Similarity 100.0%; Pred. No. 9, 6e-174;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

DB 61 KTFELDGQQLKQMDHKEAIRDAQVNGEQGRPYMTDAERVDAQYRENGFNIVYSDK 120
QY 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGWSSLLRTVSHVLRSPPELVAEI 180
DB 121 ISLNRSLPDIRHNCNCKRYLETLPNTSIIIPHNMGWSSLLRTVSHVLRSPPELVAEI 180
QY 181 VLVDSDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMIGASVATGDTVTFPLDSHC 240
DB 181 VLVDSDREHLKKPLLEDVMALEPSVRIILRTKKREGILITRMIGASVATGDTVTFPLDSHC 240
QY 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYYKRIPIP 300
DB 241 EAVNMVLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETOAGDAMRGAFDWMYYKRIPIP 300
QY 301 PELQKADPSDPFESSPVMAAGLFPAVDRKFWELGSDYDGLIHWGEQYEISFKV 353
DB 301 PELQKADPSDPFESSPVMAAGLFPAVDRKFWELGSDYDGLIHWGEQYEISFKV 353

```

```

RESULT 12
US-10-292-896-3
; Sequence 3, Application US/10292896
; Publication No. US20030186850A1
; GENERAL INFORMATION:
; APPLICANT: HASSAN, Helle
; APPLICANT: REIS, Celso A.
; APPLICANT: BENNETT, Eric P.
; APPLICANT: CLAUSEN, Henrik
; TITLE OF INVENTION: METHODS TO IDENTIFY AGENTS MODULATING FUNCTIONS OF POLYPEPTIDE G
; TITLE OF INVENTION: TRANSFERASES, PHARMACEUTICAL COMPOSITIONS COMPRISING SUCH AGENT
; FILE REFERENCE: 4305/1H154-US3
; CURRENT APPLICATION NUMBER: US/10/292, 896
; PRIOR FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/425, 204
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/DK01/00328
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/203, 331
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-292-896-3

Query Match 42.5%; Score 1392; DB 14; Length 276;
Best Local Similarity 97.3%; Pred. No. 1, 2e-126;
Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

```

```

QY 349 ISFKWMCGRMEDICSRVGHITRYKYVPYKVPAGVSLA--RNLKVAEYVMEYAEYIY 406
DB 20 LAMQVMCGRMEDICSRVGHITRYKYVPYKVPAGVSLARVRLKVAEYVMEYAEYIY 79
QY 407 QRRPEYRHLSAGDVAAYQKLRSSLNCKSPKFWMTKIAMLDLPKYPPVPEPPAAAGETIRNV 466
DB 80 QRRPEYRHLSAGDVAAYQKLRSSLNCKSPKFWMTKIAMLDLPKYPPVPEPPAAAGETIRNV 139
QY 467 GTGLCADYTKGALGSLPLRLEGCYRGRGEAAMNMVQFTFWREDIRPGDQHTKKCFDA 526
DB 140 GTGLCADYTKGALGSLPLRLEGCYRGRGEAAMNMVQFTFWREDIRPGDQHTKKCFDA 199
QY 527 ISHTSPVTLVDCHSMKGNQLMKRYKDKTYHPVSGSCMDCSBSDBHIFMNTCNPSLITQ 586
DB 200 ISHTSPVTLVDCHSMKGNQLMKRYKDKTYHPVSGSCMDCSBSDBHIFMNTCNPSLITQ 259
QY 587 WLFETHNSVTLKFNEN 603
DB 260 WLFETHNSVTLKFNEN 276

```

```

RESULT 13
US-09-795-926-27
; Sequence 27, Application US/09795926
; Patent No. US20020098486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-795-926-27

Query Match      41.8%; Score 1370; DB 9; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e-124;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNKG 157
DB 1 MTDARVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNKG 60
QY 158 WSSLRTVHSVLRSPPELVAEIVLVDDPSDRHLKKPLEDYVALPFSVRIIRTKKREG 217
DB 61 WSSLRTVHSVLRSPPELVAEIVLVDDPSDRHLKKPLEDYVALPFSVRIIRTKKREG 120
QY 218 IRTMLGASVATGDTTFFDSHCSEANVMWLPPLDRIARNRKTIVCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTTFFDSHCSEANVMWLPPLDRIARNRKTIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFSPVWAGLFAVDRKFWELGSDYD 337
DB 181 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFSPVWAGLFAVDRKFWELGSDYD 240
QY 338 GLEIWGGEQYEISFKV 353
DB 241 GLEIWGGEQYEISFKV 256

```

```

RESULT 14
US-10-364-774-27
; Sequence 27, Application US/10364774
; Patent No. US20030144497A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 269
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-364-774-27

Query Match      41.8%; Score 1370; DB 14; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.6e-124;
Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 MTDARVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNKG 157
DB 1 MTDARVDQAYRENGFNIVYSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPHNKG 60
QY 158 WSSLRTVHSVLRSPPELVAEIVLVDDPSDRHLKKPLEDYVALPFSVRIIRTKKREG 217
DB 61 WSSLRTVHSVLRSPPELVAEIVLVDDPSDRHLKKPLEDYVALPFSVRIIRTKKREG 120
QY 218 IRTMLGASVATGDTTFFDSHCSEANVMWLPPLDRIARNRKTIVCPMIDVIDHDDFRYE 277
DB 121 IRTMLGASVATGDTTFFDSHCSEANVMWLPPLDRIARNRKTIVCPMIDVIDHDDFRYE 180
QY 278 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFSPVWAGLFAVDRKFWELGSDYD 337
DB 181 TQAGDMRGAFFWEMYKRIPIPELOKADPSDPFSPVWAGLFAVDRKFWELGSDYD 240
QY 338 GLEIWGGEQYEISFKV 353
DB 241 GLEIWGGEQYEISFKV 256

```

```

RESULT 15
US-09-795-926-35
; Sequence 35, Application US/09795926
; Patent No. US20020098486A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926

```

/ CURRENT FILING DATE: 2001-02-28
/ PRIOR APPLICATION NUMBER: US 60/185,920
/ PRIOR FILING DATE: 2000-02-29
/ PRIOR APPLICATION NUMBER: US 60/186,558
/ PRIOR FILING DATE: 2000-03-02
/ PRIOR APPLICATION NUMBER: US 60/191,849
/ PRIOR FILING DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 47
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 35
/ LENGTH: 321
/ TYPE: PRT
/ ORGANISM: homo sapiens
US-09-795-926-35

Query Match 40.0%; Score 1311; DB 9; Length 321;
Best Local Similarity 99.2%; Pred. No. 1.2e-118;
Matches 251; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRRKKRLQAVLVLAALVLLPNVGLMALYREROPDGTGGGGAAPAAAGGSHSRQK 60
Db 1 MRRKKRLQAVLVLAALVLLPNVGLMALYREROPDGTGGGGAAPAAAGGSHSRQK 60
Qy 61 KTFPLGSGQKLKQWHDKEAIRDQRYNGEGGRPYMTDAERYDQAYRENGFNIVYSDK 120
Db 61 KTFPLGSGQKLKQWHDKEAIRDQRYNGEGGRPYMTDAERYDQAYRENGFNIVYSDK 120
Qy 121 ISLNRSIPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLRTVHSVLNRSPELVAEI 180
Db 121 ISLNRSIPDIRHPNCSKRYLETLPNTSIIIPHNBSGSSLRTVHSVLNRSPELVAEI 180
Qy 181 VLVDDFSREHLKKPLEDMVALPPSVRIIRTKKREGIIRTRMLGASVATGDIVITPLDSHC 240
Db 181 VLVDDFSREHLKKPLEDMVALPPSVRIIRTKKREGIIRTRMLGASVATGDIVITPLDSHC 240
Qy 241 EAVVNMVLPPLDR 253
Db 241 EAVVNMVLPPLGK 253

Search completed: December 20, 2004, 14:16:19
Job time: 151 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 14:03:56 / Search time 42 Seconds
(without alignments)
1381.398 Million cell updates/sec

Title: US-10-001-851-2

Sequence: 1 MRKERRLQAVLVLAALV.....TQOMLFHTNVTLEKFNEN 603

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392.5	42.5	684	2 T26930	hypothetical prote
2	1392	42.5	276	2 T12552	hypothetical prote
3	1125	34.3	359	2 UC4223	polypeptide N-acet
4	1115	34.0	559	2 A45987	polypeptide N-acet
5	1113	34.0	601	2 T42251	polypeptide N-acet
6	1083	33.0	624	2 T42247	polypeptide N-acet
7	1079	32.9	626	2 T42246	polypeptide N-acet
8	1078.5	32.9	623	2 T42245	probable polypepti
9	988.5	30.2	563	2 A88515	polypeptide N-acet
10	988.5	30.2	612	2 T42243	probable polypepti
11	968	29.5	571	2 T42249	polypeptide N-acet
12	949	29.0	571	2 T37405	polypeptide N-acet
13	948.5	28.9	618	2 T42248	polypeptide N-acet
14	938	28.6	562	2 T42250	polypeptide N-acet
15	934	28.5	579	2 UC5247	polypeptide N-acet
16	906	27.6	579	2 T31549	polypeptide N-acet
17	883	26.9	589	2 T42244	probable polypepti
18	801	24.4	605	2 T27397	hypothetical prote
19	632	19.3	421	2 T42252	polypeptide N-acet
20	147	4.5	1044	2 H97186	glycosyltransferas
21	139	4.2	306	2 D87531	glycosyl transfera
22	138.5	4.2	308	2 H87306	glycosyl transfera
23	136	4.1	328	2 B84263	hypothetical prote
24	135.5	4.1	312	2 S74669	hypothetical prote
25	135.5	4.1	318	2 D87506	glycosyl transfera
26	130	4.0	322	2 AC2023	hypothetical prote
27	130	4.0	787	2 H98163	hypothetical prote
28	128	3.9	470	2 C70641	hypothetical prote
29	126	3.8	313	2 A12404	hypothetical prote

30	124	3.8	362	2 G75191	dolichol-phosphate
31	124	3.8	972	2 T09595	glucuronosyltransf
32	123	3.8	334	1 G71153	hypothetical prote
33	122	3.7	316	2 AB2868	UDP-hexose transfe
34	122	3.7	316	2 F97644	UDP-hexose transfe
35	120	3.7	260	2 B90984	probable glycosyl
36	120	3.7	260	2 H85829	glycosyl transfera
37	119.5	3.6	416	2 AE1499	conserved hypochet
38	118.5	3.6	298	2 B75096	glycosyl transfera
39	118.5	3.6	416	2 AB1141	probable glucosami
40	117.5	3.6	245	2 D87307	glycosyl transfera
41	117	3.6	343	2 A12091	glucosyltransferas
42	116	3.5	251	2 A12106	hypothetical prote
43	115.5	3.5	269	2 A13123	glycosyltransferas
44	115.5	3.5	392	2 H69814	hypothetical prote
45	115.5	3.5	477	1 J50589	endo-1,4-beta-xyla

ALIGNMENTS

RESULT 1

T26930 hypothetical protein Y45P10D.3 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 09-Jul-2004

C/Accession: T26930

R/McMurray, A.

submitted to the EMBL Data Library, January 1998

A/Reference number: Z20288

A/Accession: T26930

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-684 <WIL>

A/Cross-references: UNIPROT:O45947; EMBL:AL021492; PIDD:CAAL6378.1; GSPDB:GN00022; CESP

A/Experimental source: clone Y45P10D

C/Genetic8:

A/Gene: CESP:Y45P10D.3

A/Map position: 4 112/1, 142/3, 178/3, 230/3, 289/2, 611/2

A/Intons: 61/3, 112/1, 142/3, 178/3, 230/3, 289/2, 611/2

C/Superfamily: polypeptide N-acetylglucosaminyltransferase

Query Match	Score	Length	DB 2	Score	Length	DB 2	Score	Length	DB 2
Best Local Similarity	48.8%	Pred. No. 1.6e-104;	Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;						
QY	49	PAAGQSHSRQKKTFPLGDKQKLDKWDKRAIRDAQVNGEGRPYPTDAERYDQ--	106						
DB	121	PPAALGDEALDPEKRYRGH-EKIK-WEDBAAYEKERKRGSGWGRKLPEDKEVEKEA	178						
QY	107	--AYRNGRNIVYSDKISLRSLPDRHNCNSKRYLELPNTSTIIIPFNCGSSLRRT	164						
DB	179	LSLYKANGYNAVYISDMISLRSLKIDIRHKECKMMMSAKLPVSVIFPHEBNSSTLRS	238						
QY	165	VHSVLRSPPELVAELVLDVDFSDREHLKKPLDYM---ALPSPVILRTKKRGGILRTR	221						
DB	239	VSVINRSPPELVAELVLDVDFSDREHLKKPLDYM---ALPSPVILRTKKRGGILRTR	298						
QY	222	MGASVATGDVITFLDHSCEANVMPLPLDIRAARKTIVCPMDIVDHDPRRYETQAG	281						
DB	299	QLGADATGELILFLFAHSEANVMPLPLDIRAARKTIVCPMDIVDHDPRRYETQAG	357						
QY	282	DAMRGAFDWMETTKRIPPELOKADSPDPSPVMAAGLFAVDKRFWEELGSDYGLBI	341						
DB	358	EGARSGFDMAFNKRLPLTRK--DRESPTFPFNSPVMAAGYFAISAKFWELGGYDGLDI	416						
QY	342	MGGEVYEISPKWMMCGRMEDIPCSRVGHYR-KYVPY-VVAGVSLAANLKVAEVMND	399						
DB	417	MGGEVYEISPKWMMCGRMEDIPCSRVGHYR-KYVPY-VVAGVSLAANLKVAEVMND	476						
QY	400	EVAEYIYQRRPEYRHLSADVAVQKLRSLNCKSPKFWMTKIAMDLPKFPVPEBPAA	459						
DB	477	DYKETLYKRRPVGNNADGDLKLMKGIKRLCKSPDMWKEIAFDQDKYYPVPEBPAA	536						

QY 460 MGEIRVGTGLCADTKHAGLSPLRLEGVCR-----GRGEAAMNNMVFPTFWREDIRPGD 515
 |||||
 Db 537 EGEIRVGTGTCIDTQKEQNGRFGRLKCTSDKDGGE-----QDLRLTRHMDIRP-- 588
 |||||
 QY 516 PQTAKFCFPAIAHNT--SPPTLYDCHSMKGNQIMKTR-KKTLVHPVSGSCMDCSESDHR 572
 |||||
 Db 589 --KGRKICFCDCSTSVKAPVILPDCCHSMKGNQIMKTRVAKQIYHPISGCLTRADENGKG 646
 |||||
 QY 573 -FMNTCNPSLTLQOVMFEHTNSTVLEKFNRN 603
 |||||
 Db 647 FLHMKKCDSSDLQKAMQTVNDELLETROAN 678
 |||||

RESULT 2

hypothetical protein DKFZp586h0623.1 - human (fragments)
 T12552
 C/Spectrum: Homo sapiens (man)
 C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
 C/Accession: T12552
 R/Author: W. J. Mewes, H. W. J. Gaassenhuber, J. J. Wiemann, S.
 Submitted to the Protein Sequence Database, June 1999
 A/Reference number: Z17527
 A/Accession: T12552
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-150/151-276 <ANS>
 A/Cross-references: UNIPROT:O9Y4M4; EMBL:AL096739
 A/Experimental source: adult uterus; clone DKFZp586h0623
 A/Note: the cDNA sequence contains a -1 frameshift near codon 150
 C/genetics:
 A/Note: DKFZp586h0623.1
 C/Superfamily: polypeptide N-acetylgalactosaminyltransferase

Query Match 42.5%; Score 1392; DB 2; Length 276;
 Best Local Similarity 97.3%; Pred. No. 4.6e-105;

Matches 250; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 349 ISFKVWCGRMEDICSRVGHVYKRYVPAVAGVLA--RNILKRVAEVMDVAYEYI 406
 |||||
 Db 20 LAMQVWCGRMEDICSRVGHVYKRYVPAVAGVLA--RNILKRVAEVMDVAYEYI 79
 |||||
 QY 407 QRRPVRHLSAGVAVVOKLRSSSLNCKSFEMTKIAMDLPRKYVPVPEPAAWGEIRNY 466
 |||||
 Db 80 QRRPVRHLSAGVAVVOKLRSSSLNCKSFEMTKIAMDLPRKYVPVPEPAAWGEIRNY 139
 |||||
 QY 467 GTGLCADTKHAGLSPLRLEGVCRGEAAMNNMVFPTFWREDIRPGDQHTKKCFPA 526
 |||||
 Db 140 GTGLCADTKHAGLSPLRLEGVCRGEAAMNNMVFPTFWREDIRPGDQHTKKCFPA 199
 |||||
 QY 527 ISHTSPVTLVYDCHSMKGNQIMKTRKDTLYHPVSGSCMDCSESDHRIFMNTCNPSLTLQ 586
 |||||
 Db 200 ISHTSPVTLVYDCHSMKGNQIMKTRKDTLYHPVSGSCMDCSESDHRIFMNTCNPSLTLQ 259
 |||||
 QY 587 MFEHTNSTVLEKFNRN 603
 |||||
 Db 260 MFEHTNSTVLEKFNRN 276
 |||||

RESULT 3

polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - human

N/Alternate names: GalNAC-transferase
 C/Species: Homo sapiens (man)
 C/Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004

C/Accession: J04223; J17404
 R/Author: J. A. J. Mayhew, J. N. Baker, C. A. J. Thomsen, D. R. J. Homa, F. L. J. Elhammer, A. P.

J. Biochem. 118, 568-574, 1995
 A/Title: cDNA cloning, expression, and chromosomal localization of a human UDP-GalNAC:

A/Reference number: J04223; MIMD:96115928; PMID:8690719

A/Accession: J04223
 A/Molecule type: mRNA
 A/Residues: 1-559 <MEU>

A/Cross-references: UNIPROT:Q10472
 A/Experimental source: salivary gland
 A/Note: The authors translated the codon AAT for residue 264 as Asp
 R/White, T. J. Bennett, R. P. Paul, E. J. Takio, K. J. Sorensen, T. J. Bonding, N. J. Clausen, H.
 J. Biol. Chem. 270, 24156-24165, 1995
 A/Title: Purification and cDNA cloning of a human UDP-GalNAC:polypeptide N-Acetylgalactose
 A/Reference number: J17404; MIMD:96025800; PMID:7592619
 A/Accession: J17404
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-559 <RES>
 A/Cross-references: EMBL:X85018; NID:9971458; PIDN:CA59380.1; PID:9971459
 A/Comment: This enzyme catalyzes the initial reaction in O-linked (mucin type) oligosac
 ccharide pathway.
 C/Genetics:
 A/Map position: 18
 C/Superfamily: polypeptide N-acetylgalactosaminyltransferase
 C/Keywords: Chromosomal protein; glycoprotein; glycosyltransferase; hexosyltransferase;
 F/3-28/Domain: transmembrane #status predicted <TMM>
 F/29-559/Domain: endoplasmic reticulum lumenal #status predicted <LUM>
 F/35,141,541,552/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/117,118,288/Binding site: carbohydrate (Thr) (covalent) #status predicted
 F/119/Binding site: carbohydrate (Ser) (covalent) #status predicted

Query Match 34.3%; Score 1125; DB 2; Length 559;
 Best Local Similarity 44.0%; Pred. No. 5.6e-83;

Matches 232; Conservative 87; Mismatches 170; Indels 38; Gaps 16;

QY 88 GNGEQR--YEMTDKERDQAVRENGEVIYSDKLSLRSLPDRIRHPCNSKRYLETLP 145
 |||||
 Db 57 GPEMKRPVPIPEDEKKEKMFKNQFNLMSSEMIALNRSJLDVLECKTKVYPDNLP 116
 |||||
 QY 146 NTSIIIPFNEGSSSLRTVHSLNRSPELVAEIVLVDPFDRHLEKPLBDM-ALRP 204
 |||||
 Db 117 TTSVAVFVFNEMASTLKTAVHVINRSPHMIETIVDABERDLPKPLSYKGLKV 176
 |||||
 QY 205 SVRIITKRRBGLIRTKMGASVAGTDTITFLDSHCANVNLPLLDRIARNKRTIVCP 264
 |||||
 Db 177 PVHVMQRSGILIRALTKGAIVSKQVITFLDAHCECTVGMLEPLARIKIDRRIVCP 236
 |||||
 QY 265 MLDVIDHDPRFETQAG-DAMGAPDMEVYKRIPIRP--ELQKADBPDPESPVMAG 320
 |||||
 Db 237 IIDVSDDTFEY--MAGSDMTYGGFNWKLNFMYVPVQREMRKRDRLPVRTPTMAG 294
 |||||
 QY 321 LFAVDKFWELVGGYVPGLEIWGGEQYELSFVKWCGRMEDICSRVGHVYKRYVPA 380
 |||||
 Db 295 LPSIDNDYQELGTIDAGNDINGENLEISFIMQCGTLEIVTSGHGVHVRKATPIYF 354
 |||||
 QY 381 PAVGS--LARNILKRVAEVMDVAYEYIYQRPDEYHLSAGVAVVOKLRSSSLNCKSF 438
 |||||
 Db 355 PGDTGQIINKNRRRLAEVMDVDFKPFYIISGVTVKVDGDISRVGLAHKLQCKPFSWY 414
 |||||
 QY 439 MKRIAND--LPKFPVPEPAAWGEIRVAVGGLCADTKHAGLSPLRLEGVCRGEA 496
 |||||
 Db 415 LENIYPSQIPRHY-----PSLGEIRNVEVTEVQCDNNARKENKVGIFNC-HGNG-- 463
 |||||
 QY 497 MNMNVPTFWREDIRPGDQHTKKCFPAISHTSPVTLVYDCHSMKGNQIMKTRKDT 555
 |||||
 Db 464 --GNQVFSYTKANKETITD-----LCDVSLNGVPMVLMKCHNLKGNQIMVDPVKT 515
 |||||
 QY 556 YHPVSGSCMD-CSESDHRI-FMNTCNPSLTLQOVMFEHTNSTVLEKFN 600
 |||||
 Db 516 QHVNNSQCUDKATEEDSQVPSIRDCN-GSRSQOMLIR--NVTLPEIF 559
 |||||

RESULT 4

polypeptide N-acetylgalactosaminyltransferase (EC 2.4.1.41) - bovine

C/Species: Bos primigenius taurus (cattle)
 C/Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004

C/Accession: A45987; A48530
 R/Homa, F. L. J. Hollander, T. J. Lehman, D. J. J. Thomsen, D. R. J. Elhammer, A. P.

J. Biol. Chem. 268, 12609-12616, 1993

Query Match	34.0%;	Score 1115;	DB 2;	Length 559;
Best Local Similarity	43.6%;	Pred. No. 3.6e-82;		
Matches 230; Conservative	89;	Mismatches 170;	Indels 38;	Gaps 16;

RESULT 5

Query Match 34.0%, Score 1113, DB 2, Length 601,
Best Local Similarity 42.0%, Pred. No. 5.8e-82;
Matches 226, Conservative 90, Mismatches 186, Indels 36, Gaps 13;

A:Molecule type: mRNA
A:Residues: 1-601 <HAG>
A:Cross-references: UNIPROT:O61397, EMBL:AF031841, NID:G30472202, PIDD:AA31677.1, PIDD:G30472202
C:Genetics: gly-7
C:Superfamily: polypeptide N-acetylglucosaminyltransferase
C:Keywords: glycosyltransferase, hexosyltransferase

Query Match 33.0%, Score 1083, DB 2, Length 624,
Best Local Similarity 45.0%, Pred. No. 1.7e-79;
Matches 226, Conservative 76, Mismatches 166, Indels 34, Gaps 13;

```

0Y 103 VVDQAVRENGFPIYVSDXISLNRSLPDIRHPNQNSKXLETLPTMPSIIIPPHNGWSL 162
0Y 133 KYDKMMLNARQVASDMISVHRTIPTNIDAECKTEKNEUPRTSVIICHANMAWSL 192
0Y 163 RTVHSLNRSPEELVAELIVLDDFSDREHLKPLEDYMAF-PSVRIIARTKKEGLIRTR 221
Db 193 RTVHSLVLRTPDHLLEEVVLVDDFSDMDHTRPLEEYMSQFGAKVILIRMEKREGILIRAR 252
0Y 222 MLGASVARGDITFLDSCHEANVMMLPBLIRIANRRTIICPMIDVIDHDDEFY-ETQA 280
Db 253 LRGAANVAIGEVLITYDSCHECEMGEMLBIRIKDPTTVVCPIYDIVDNDTEYHNSKA 312
0Y 281 GDAMGAFDWMYKRIPIPELOK-ADPSDFEESPVMAGSLPAVDKMFELGGYDPG 338
Db 313 YFTSVGSGDMQIOFNWHSI PERDRKNRTRPIDPVASPTMAGLSIDIXEYFEKLGTDPG 372
0Y 339 LEIMSGEYELISFKYMMGGMRMEDI PCSRVGHIIYKRYPVKPVAGS-LAENLRVAVW 397
Db 373 FDIIMGELNLEISFKIMCGGTLEIVPCSHVGVFFKRS PYKMRGVANLKNNSIRLAEW 432
0Y 398 MDEVAELIYORRPEYRNLASGDVAQOKLRSLSLCKSEKFMETKIAMDLPKYPVPVEBPA 457
Db 433 LDDYXTYYEYRINNDG-DFGDISRKKLRBDLGCKSPKWLNDI---YPLFVPGE-S 486
0Y 458 AANGSIRAVGICLCAVDTKHGLSGPLREG----CYARGBEAANNQVFTFWREDIR 512
Db 487 VAKGSLRNAQSQCLDS---AVGEVEVENKATTPYPCHEQGNQYU-----MLSKDGIR 537
0Y 513 PGDPQHTKKCFEDAISHTSPVTLYDCHSMKGNQLMKYRKD-TLYHPVSGSCMDSESDH 571
Db 538 RDES-----CYDYG--SDVMVFPCHGMKNGQEBRYNHDTGRLOHVASQCLMTXOGA 589
0Y 572 RIFMNTCPSSILTOQMLPEHTN 593
Db 590 KLEWACQYDDPYOHMKFEYN 611

```

RESULT 7

T42246

polypeptide N-acetyl-galactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis elegans*

C|Species: *Caenorhabditis elegans*

C|Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C|Accession: T42246

C|Hagen, F.K.; Nehrkne, K.

J. Biol. Chem. 273, 8268-8277, 1998

A|Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polypeptide

A|Reference number: Z22126; MUID:98192620; PMID:9525933

A|Accession: T42246

A|Status: preliminary; translated from GB/EMBL/DBJ

A|Molecule type: mRNA

A|Residues: 1-626 <HAG>

A|Cross-references: UNIPROT:O61392; EMBL:AF031836; NID:G3047192; PIDN:AC13672.1; PID:G3047192

C|Genetics:

A|Gene: gly-5

C|Superfamily: polypeptide N-acetyl-galactosaminyltransferase

C|Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1079; DB 2; Length 626;

Best Local Similarity 44.9%; Pred. No. 3.5e-79;

Matches 227; Conservative 75; Mismatches 164; Indels 40; Gaps 14;

QY 103 RYDQAARENGFNIVYSDKISLNRSPLDRIHNCNSKRLLETLPNNSIITPPNEGSSLL 162

DB 133 KYDKGMLNNFNQASDMISVHRTLPTNIDAECKTEKNEINPRTSVIICFHNAAVSLL 192

QY 163 RTVSHVLTNRSPBELVAEIVLVDFSDREHLKKPLLEDYNAF-PSYRIITKKREGILTR 221

DB 193 RTVSHVLTNRSPDLHEEVLVLDPSDMOHTKRPLEEYWSQFGKVKILFEMEREGILRR 252

QY 222 MGASVATGTVITFLDSHCENANVMVPLRLRIARNKTIYCPMLDIVIDHDPRY-ETQA 280

DB 253 LKGAATAATGVLVLTLDSHCEGMEGMEPLDRIKDDPTTVACPVLDIVDNTFEYHNSA 312

```

0Y      28  GDAMRGAPMBWBYKKRIP1PELQK--ADSPSPFSSPVMAGLPAVDRKPMWELAGYDPG 338
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      313 YFSSVGFDMGLDFNMHSTIPERDKORRPRPIDVPARSPIMAGGLSIDKEYFEKGLTYDPG 372
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
0Y      339 LEIWGBOYE1SEFKVMCGGMEDIPCSRGVGH1YKVPYKVPAGVS-LARNLKRVAEW 397
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      373 FDIIMGNE1ELSKT1MMGCGTLEIYPCGHVGHVFPKRSPYKMRGTGVATLKNSIRLAEW 422
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
0Y      398 MDEYAEY1YORBEYRHLASGVAVAVOKLRSSLNCKSEFKMFTYIANDLPEKY1PVPBP 457
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      433 LDYKTYTYER1NNQK--DFGDISRRKKLRDELCKSEFKW1LDNI--YELFVPGS--S 486
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
0Y      458 AANGELR1NVTGLCADTKHGLGSPRLLEGCYRGGEGEAAMNMVYFTT-----WR 508
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      487 VANGEVYNSAV-----QBARCLDCCVNGHE---NDRPVGYTCHGGGNGYWM 531
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
0Y      509 EDLRPGDPQTKKKECFDA1SHTSPTLVLDCHSMKGNQ1AMKRXOK-PLYHPVSGSCMDCS 567
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      532 1S-KDGGIRDES-CVDYAG--SDVMVPPCHGMKGNOEMRYNRHDTGRLOHVAHSOKCLGMT 587
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
0Y      568 ESDR1FPMNTCNPSLS1QOOLPEHTN 593
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      588 KDGAKLBNACQYD1DYQHMKFEYN 613
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 8

T42245

probable polypeptide N-acetylglactosaminyltransferase (EC 2.4.1.41) - *Caenorhabditis el*

C:Species: *Caenorhabditis elegans*

C:Date: 03-Dec-1999 #sequence_rev:revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T42245

R:Hagen, F.K.; Nehrkke, K.

J: Biol. Chem. 273, 8268-8277, 1998

A:Title: cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine:Polypept

A:Reference number: Z22126; NCID:98192620; PMID:9555933

A:Accession: T42245

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-623 <HAG>

C:Cross-references: UNIPROT:O61391; EMBL:AF031335; NID:g3047190; PIDN:AA013671.1; PID:g3

C:Genetics:

A:Gene: gly-5

C:Superfamily: polypeptide N-acetylglactosaminyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 32.9%; Score 1078.5; DB 2; Length 623;

Best Local Similarity 45.3%; Pred. No. 3.8e-79;

Matches 226; Conservative 74; Mismatches 170; Indels 29; Gaps 13;

QY 103 RVDQAVRENGFNIVYSDKISLNSLPDIRHPNCSKRYLETLPNTSIIIPHNEGWSSLL 162

DB 133 KYDKGMLNNAFQOYADMSIVHRTLPNTIDAECKTEKYENELPRTSVIICHNEAMSVLL 192

QY 163 RTVHSLVANSPELVVAEIVLVDDFSDREHLKRPLEDYMALF-PASVIRLRTKKREGILRTR 221

DB 193 RIVHSLVLETPPHLEBEVLVDDFSDMDHTKRPLEBYMSQPGKVKILIMRKREGLIAR 252

QY 222 MLAGSVAATGDTVTFLDSDCEANVWMLPELLDIRIARNKTYCPMIDVIDHDPRY-ETQA 280

DB 253 LRGAVAATEBEVLTLDSDHCECMEGMMBPELLDIRIKRDTPTVCPIYDIVDDMTFFEYHNSKA 312

QY 281 GDAMGAPFMEMVYKRIPIPELQK--ADPSDPESPVMMAGLFAVADRKWMELGQYBG 338

DB 313 YFTSVGGFWMGLQFNMHNSIPERDRKRTIRPIDRSPVTMMAGLSIDKEVFKLGSTYDGG 372

QY 339 LEIMGGEQYEISFKVMMCGGRMEDIPCSRVGHYIRKVVYPYKVPAGVS-LAENLRKVAEVM 397

DB 373 FPIWGGENLELSPFKIMWGCGTLEIVPCSHVGHVFRKKSPLYKWRGVNVILKENSIRLAEVM 432

QY 398 MDEYAEYIQRREPEYHNSAGDVAVQKRLSSLNCKSPKKFMFTLIANDLPKFFYPPEVERPA 457

DB 433 LDDYKTYTYERINNQGL-DFGDISSRKRLREDLGCKSPKFWYLDNI---YPELFLPVG--S 486

T42249
 poly(ADP-ribose) polymerase (EC 2.4.1.1) 6b - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T42249
 R/Hagen, F.K.; Nehrkke, K.
 J. Biol. Chem. 273, 8268-8277, 1998
 A/Title: CDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:poly(ADP-ribose) transferase
 A/Reference number: 222126; PMID:98192620; PMID:9525933
 A/Accession: T42249
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-617 <HAG>
 A/Cross-references: UNIPROT:O61395; EMBL:AF031839; NID:93047198; PID:AA013675.1; PID:93
 A/Genes: gly-6
 C/Superfamily: poly(ADP-ribose) transferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.5%; Score 968; DB 2; Length 617;
 Best Local Similarity 39.3%; Pred. No. 3,4e-70;

Matches 216; Conservative 96; Mismatches 162; Indels 76; Gaps 22;

QY 73 DHDKAIRRDAQRVNGEGRPYMTDAERV-DQAYRENGFNIVSDKISLNRSLPDIR 131
 DB 97 DW-----GGGAGVSHLPTEQOKLADSTFAVQNFILVSDGISVRSLSPEIR 143
 QY 132 HPCNSKRYLETLPNTSIIIPFNEGSSLLRTVSHVLSNRPPELVAEIVLVDPDREH 191
 DB 144 KPSGRNMTYFDNPTTSVIVHNEAVSTLRTVSWTIDSPKLEIILVDPDREH 203
 QY 192 LKRP-LEDYVALRPS-VRIILRTKKRGLIRTRMIGASVATGDTYTFDLSHCANVWLP 249
 DB 204 LRPTLDTTLKPLPTIKIRSKERGLIRAMMGAGAGDVLTFDLSHCETKMLER 263
 QY 250 LDRIRANKRTIVCPMIDVIDHDDFRYETQAGAMRGAPFWEYIKRIPPELQK--A 306
 DB 264 LTRIKLRKAVPCPVLDIINDTFOYQ-KGIEMFEGGFWMNQFRYWGPTMAKQHL 322
 QY 307 DPEDPESPVMAGLFAVDKMFELGVDGLEINGGEOYETSPKVMCGMEDIKPS 366
 DB 323 DPGPTESPTMAGGLSINNYEELGSDYDGMIDGEMLEMSFRMQCGRVELLP 382
 QY 367 RVGHIRKRYVYKVP--AGVSLARMLKVAEYVMDYAEIYQRPPE-VRHLSAGDAV 422
 DB 383 HGVHVRKSSPHFPFGSSGKVLNTNLRAVEYVMDWKHYFKIAPQARMSSIDVSE 442
 QY 423 QKKLRSSLNCKSKFMFTKIAMDLPKFYPPVPPAAWGEIRVVGGLC---ADT--- 474
 DB 443 RVELRKLNCKSKFMYLQNVFOD--HFLP---TFLDRFGMSN--SNYCTAFRPGDTPK 495
 QY 475 KHGALGPLRLBECVGRGAANNMNVFTWRBDRPDPQHTKKFCFDALS--HTS- 531
 DB 496 NHRLLSP---CTMG---FDLMQWMLYTDGRRIRIDE---HLCLSVQLHTTS 540
 QY 532 --PVTLYDCHSMKNQMLKTR-KDKTLVHPVSGSCMDCSBSDRIF-----MNT 577
 DB 541 DMKIQLEKCGF-DTEVWDFKPKIGRFQNRKTKGLCLASP-----IDPTKDEENPPIVQ 595
 QY 578 CNPSSLTQW 587
 DB 596 CRSSNDROW 605

RESULT 12
 T42405
 poly(ADP-ribose) polymerase (EC 2.4.1.1) - human
 C/Species: Homo sapiens (man)
 C/Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
 C/Accession: T42405
 R/Hagen, F.K.; Nehrkke, K.; Paul, E.; Takio, K.; Srensen, T.; Bonding, N.; Clausen, H.
 J. Biol. Chem. 270, 24156-24165, 1995
 A/Title: Purification and CDNA cloning of a human UDP-GalNAc:poly(ADP-ribose) transferase

A/Reference number: T42405; PMID:7592619
 A/Accession: T42405
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-571 <RES>
 A/Cross-references: UNIPROT:Q10471; EMBL:X85019; NID:99711460; PID:CAA59381.1; PID:997114
 A/Genes: GALNT2; GALNAC
 A/Map position: 16q24-16q24
 C/Superfamily: poly(ADP-ribose) transferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 29.0%; Score 949; DB 2; Length 571;
 Best Local Similarity 33.5%; Pred. No. 1.1e-68;

Matches 221; Conservative 103; Mismatches 170; Indels 166; Gaps 21;

QY 1 MRRKEKLLQAVLVALLVLLPNVGMALYREKQDPGSGGAAPACQSHSROK 60
 DB 1 MRRSRMLLCFAFLWLTGIA-----YMY-----SGGSLAGAGGGA----- 39
 QY 61 KTFPLGSGKLDMDHKEAIRRDAQRVNGEG-----GR-PYMTDAERV----- 104
 DB 40 -----GRK-EDWNEIDPTKKDLHNSGEEKAGSMETLPPGKRVMPDNPQAYVGTW 91
 QY 105 ----DQAYRENGFNIVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIIPFNEGWS 160
 DB 92 VNSGDPYRANKFNQVSESKLMDRAIPTRIDQCKRMARDLPATSVITFHNBARA 151
 QY 161 LRTVSHVLSNRPPELVAEIVLVDPDREHKKPLED--VVALPSPVILRTKKREGI 218
 DB 152 LRTVSVLVKSPPHLIKELIILVDDYSNDP-----EDGALLGKIEKVALVLRNDREGIM 205
 QY 219 RTRMIGASVATGDTYTFDLSHCANVWLPPLDIRARKRTIVCPMIDVIDHDDFRYET 278
 DB 206 RSRVRGADAQAKVLFPLDSHCENHMLEPLERVAEDRTVSPITDIVINDNFQYVG 265
 QY 279 QAGDARCAFDEMYKRIPIPELQK---DPSPEESPVMAGLFAVDKMFELGYS 335
 DB 266 ASAD-LKGFMDNVNLYKMYMPBEQRSRQGNVPAPIKTPMAGLFVWDKRYFELGK 324
 QY 336 DPELEIWSGEQYEIFSKVMCGRMEDIPCSRVGHYIKRYVYKVP--AGVSLARMLK 393
 DB 325 DMMDVWCGENLEISFRVWQCGSLLEIIPCSRVGHYFRKHQHYTFPGSGGTAFANTRBA 384
 QY 394 AEVMDVEYAEIYQRPPEVRHLSAGDAVQKKLRSSLNCKSKFMFTKIAMDLPKFY 453
 DB 385 AEVMDVEYKPYFAVPSARVPYGNIGSLRLKRLSKPFRMYLENV-----YPEL 437
 QY 454 EPPAAWGEIRVVGGLCADTGHGALGSPRLLEGCVGRGAANNMNVFTWRBDRP 513
 DB 438 RVPDH-----QDIAPGL----- 450
 QY 514 GPDQHTKFCFPAISHTSP--VTLVYDCHSMKNQMLKTRKDKTLVH-----PVSGS 562
 DB 451 --QGGTN--CLDTLGHFADGVGVVECHNAGNQEALTKERSVGHMDCLTVVDRAPS 506
 QY 563 CM---DCSSDR-----IMNNTCNPSSLTQW 589
 DB 507 LKLGCRENDSRQKWEQIEGNSKLHVGSNCLDSRTAKSGGLVEYVGP-ALSOQMKF 565

RESULT 13
 T42248
 poly(ADP-ribose) polymerase (EC 2.4.1.1) - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T42248; T23138
 R/Hagen, F.K.; Nehrkke, K.
 J. Biol. Chem. 273, 8268-8277, 1998
 A/Title: CDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:poly(ADP-ribose) transferase
 A/Reference number: 222126; PMID:98192620; PMID:9525933
 A/Accession: T42248

A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-618 <HAG>
 A/Cross-references: UNIPROT:O61394; EMBL:AF031838; NID:G3047196; PIDN:AAIC3674.1; PID:93
 R;Barlow, K.
 Submitted to the EMBL Data Library, June 1998
 A/Accession number: T23138
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-618 <HAG>
 A/Cross-references: EMBL:AL024499; PIDN:CAA19707.1; GSPDB:GNO0021; CESP:H38K22.5
 A/Experimental source: clone H38K22
 C/Genetics:
 A/Map position: 3
 A/Intons: 42/3; 64/3; 154/1; 202/1; 368/3; 430/1; 477/3; 513/3; 563/3; 604/3
 A/Note: g1y-6
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 28.9%; Score 948.5; DB 2; Length 618;
 Best Local Similarity 38.3%; Pred. No. 1.3e-68;
 Matches 211; Conservative 100; Mismatches 163; Indels 77; Gaps 21;

73 DMHDKAIRRDAGRVNGSGRPRYPMTDARV--DQAYRENGFNIVYSDKISLRSLPDIR 131
 97 DW-----GEGGAGVSHLTPREQKLDSTFVAVNQFNLVSDGISVRSRLPEIR 143
 132 HPNCNSKRYLETLPNTSIIIPFNHGWSSLLRTVSHVLSNRSPELVAEIYLVDFSDREH 191
 144 KPSCRMNTYPDNLPPTSIIIVYHNEASTLIRTVMSYIDSFKELKEIILVDDFSDREF 203
 192 LKRP-LEDYALPPS-VRIIRTKRGLIRTRMLGASVATGDIYITLDSHCANVWMLPP 249
 204 LRPTLDTTLKPLPTDIKIRSKERVGLIRARMGAQOAGDVLTLFDSHCCTKGMLEP 263
 250 LDRIRARNRTIYCPMIDVIDHDPRYETQAGDAMGAPDMEMVYKRIPIPELQK---A 306
 264 LTRIRKLRKAVPCPIVDIINDNTFOYQ-KGIEMFGGFRNMLQFRWYGMPTMAKQHL 322
 307 DPSPDPSPVWAGLFAVDRKFMELGVDPLGELWGEQYEISFKVMCGGMEDIPCS 366
 323 DPTGPISPTMAGLSPINNNYFELGEYDPCMDINGENLEMSFRWOCGAVELLPCS 382
 367 RVGHIRKRYVPYKVP--AGVSLARNLKRYAEVWMEYAEIYQRPPE-YRHLISADYAV 422
 383 HVGHVFRKSSPHDFPKSSGKVLNTMLRVAEVMMDWMGKYFYKIAPOAHMRMSIDVSE 442
 423 OKTLRSSLNCKSFPMFTKIAMDLKPYRVEPPPAAMGRI-RNVTGLCADTKHGLGS 481
 443 RVBLRKLKLNCKSFKMYLQNVFOQ--HFLP--TPLDRFGMTSSNSVSC-----LAW 490
 482 PLRLGEG-----CVNGRGEAAMNNQVFTFWREBIRPGDPOHTKKPFDAIS--HTS 531
 491 TLASSGKTASTADCLK-----IFHKQMLWLYGSDRIRIRDE-----HLCISVQLLHTT 540
 532 ---PVTLYDCHSMKQNLWKYR-KDKTLVHAPVSSCMDCSBSDRIF-----NM 576
 541 SDWKIQLKECAGP-DTEYVDFKPKRIGFQNRKTKGLCLASPD---IDPPTKDFENRPIYQ 595
 577 TCNPPSLITQOW 587
 596 KCRSSNDRQW 606

RESULT 14

T42250
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) 6c - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T42250
 R;Hagen, F.K.; Nehrke, K.

J. Biol. Chem. 273, 8268-8277, 1998
 A/Title: CDNA cloning and expression of a family of UDP-N-acetyl-D-glucosamine:Polypep
 A/Reference number: Z22126; MIDID:98192620; PMID:9525533
 A/Accession: T42250
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-562 <HAG>
 A/Cross-references: UNIPROT:O61396; EMBL:AF031840; NID:G3047200; PIDN:AAIC3676.1; PID:93
 C/Genetics:
 A/Map position: 3
 A/Intons: 42/3; 64/3; 154/1; 202/1; 368/3; 430/1; 477/3; 513/3; 563/3; 604/3
 A/Note: g1y-6
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 28.6%; Score 938; DB 2; Length 562;
 Best Local Similarity 47.5%; Pred. No. 8.1e-68;
 Matches 182; Conservative 74; Mismatches 103; Indels 24; Gaps 8;

73 DMHDKAIRRDAGRVNGSGRPRYPMTDARV--DQAYRENGFNIVYSDKISLRSLPDIR 131
 97 DW-----GEGGAGVSHLTPREQKLDSTFVAVNQFNLVSDGISVRSRLPEIR 143
 132 HPNCNSKRYLETLPNTSIIIPFNHGWSSLLRTVSHVLSNRSPELVAEIYLVDFSDREH 191
 144 KPSCRMNTYPDNLPPTSIIIVYHNEASTLIRTVMSYIDSFKELKEIILVDDFSDREF 203
 192 LKRP-LEDYALPPS-VRIIRTKRGLIRTRMLGASVATGDIYITLDSHCANVWMLPP 249
 204 LRPTLDTTLKPLPTDIKIRSKERVGLIRARMGAQOAGDVLTLFDSHCCTKGMLEP 263
 250 LDRIRARNRTIYCPMIDVIDHDPRYETQAGDAMGAPDMEMVYKRIPIPELQK---A 306
 264 LTRIRKLRKAVPCPIVDIINDNTFOYQ-KGIEMFGGFRNMLQFRWYGMPTMAKQHL 322
 307 DPSPDPSPVWAGLFAVDRKFMELGVDPLGELWGEQYEISFKVMCGGMEDIPCS 366
 323 DPTGPISPTMAGLSPINNNYFELGEYDPCMDINGENLEMSFRWOCGAVELLPCS 382
 367 RVGHIRKRYVPYKVP--AGVSLARNLKRYAEVWMEYAEIYQRPPE-YRHLISADYAV 422
 383 HVGHVFRKSSPHDFPKSSGKVLNTMLRVAEVMMDWMGKYFYKIAPOAHMRMSIDVSE 442
 423 OKTLRSSLNCKSFPMFTKIAMDLKPYRVEPPPAAMGRI-RNVTGLCADTKHGLGS 481
 443 RVBLRKLKLNCKSFKMYLQNVFOQ--HFLP--TPLDRFGMTSSNSVSC-----LAW 490

RESULT 15

JC5247
 polypeptide N-acetylglucosaminyltransferase (EC 2.4.1.41) T3 - mouse
 N/Alternate names: protein-UDP acetylglucosaminyltransferase
 C/Species: Mus musculus (house mouse)
 C/Date: 13-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: JC5247
 R;Zera, J.; Hagen, F.K.; Ten Hagen, K.G.; Van Wuyckhuysse, B.C.; Tabak, L.A.
 Biochem. Biophys. Res. Commun. 228, 38-44, 1996
 A/Title: Cloning and expression of mouse UDP-GalNAc:polypeptide N-acetylglucosaminylt
 A/Reference number: JC5247; MIDID:97069650; PMID:8912633
 A/Accession: JC5247
 A/Molecule type: mRNA
 A/Residues: 1-633 <RAR>
 A/Cross-references: UNIPROT:P70419; GB:U70538; NID:G1575722; PIDN:AA09579.1; PID:G1575
 A/Experimental source: testis
 C/Comment: This enzyme catalyzes the formation of N-acetyl-D-glucosamine alpha-O-Ser/
 itating O-glycosylation of serine and threonine residues on an array of glycoproteins.
 C/Superfamily: polypeptide N-acetylglucosaminyltransferase
 C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 28.5%; Score 934; DB 2; Length 633;

Best Local Similarity 39.2%; Pred. No. 2e-67;
 Matches 206; Conservative 98; Mismatches 179; Indels 42; Gaps 18;
 90 GEGRPYPMTDARVQAVRENG-----FNIVYSDKISLRSLPDIRHPNCNSKRY--L 141

Db 122 GASGKPFKITHLSPBQCKEKEGRTYKCFNAPASDRISLHRDLGPDTRPEECIEQKFKRC 181
QY 142 ETLPTNTSIIIPFNEGSSILRTVHSVLYNSPPELVAEIYVDDFSDREHLKKPLSDYMA 201
Db 182 PPLPTTSVIVFNNANSTLRTVHSVLYSSPAILEKILTVDDASVDDYLHEKLEBYIK 241
QY 202 LPPSVRILRTYKKEGLIRTMLGASVATGDTVITFLDSHCANVNWLPPLIDRIARNRKT 261
Db 242 QFETIVKIVRQEKGLITARLGAAVATAETLTFLDHCECFYGMLEPILARIAYNTAV 301
QY 262 VCPMIDVITHDDRRYETQA--GDAMRGATDWMNTY--KRIPIPELQKADPSDPFSPV 316
Db 302 VSPDIASIDLNTFEFNNKPSPYGSHNNKGNPDWSLSFGWESLPDHEKORRDETYPIKTP 361
QY 317 MAGGLFVADRKFMELGGYDPGLIEMGEGYEISFKVMGCGRMEDIPGSRVGHYRKVY 376
Db 362 FAGGLFSISKKYEHIGSYDEMEIMGENIEMSFVWQCGQLEIWPCSVGVGHVFRSKS 421
QY 377 PYKVPAGVS-LARNLKRVAEVMNDEVAEYIYORRPE---YRHLASGDVAVQKKRSSLN 431
Db 422 PHTFPKGTQVIANNQVRLAEVMMDEYKEIFYRNTDAKIVKQKSGDLSKREIKKRLQ 481
QY 432 CKSGFKPMTKIANDLPKFYPPVBPAAAMEIRNVGTGLCADT-KHGAUGSPRLBGCVR 490
Db 482 CKNFTWYLNNTI--YPEAYVPDLNPVIS-GYIKSVGQPLCLDVGENNQGKPLILYTC-H 536
QY 491 GRGEAAMNNMNVFTFTWRREDIRPGDPQHT--KKCFDAISHTSPV--TLYDCHS--MKN 544
Db 537 GLG-----GNQYFEYSAQRER-----HNIOKEICLHATOGVVQLKACVTKGHTIAPGE 586
QY 545 QLMKRYRKDKTLVHPVSGSCMDCSESDHRIFMNTCNPSLTQOWLF 589
Db 587 QIWEIRRDQLLYNPLPFMCILS-SNGEHPNLV-PCDATDILLQKWIF 629

Search completed: December 20, 2004, 14:13:49
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2004, 13:55:48 ; Search time 197 Seconds

(without alignments)
1761.172 Million cell updates/sec

Title: US-10-001-851-2

Perfect score: 3278

Sequence: 1 MRKKEKRLQNALVALVALV.....TQGMLEPHNINSTVLEKFNEN 603

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3278	100.0	603	1 GL10_HUMAN	Q868r1 h polypepti
2	3157	96.3	603	1 GL10_HUMAN	Q925r7 r polypepti
3	3138	95.7	634	2 BAD21405	Bad21405 mus muscu
4	3137	95.7	603	1 GL10_MOUSE	Q6p9e7 m polypepti
5	3137	95.7	603	2 AAH60617	Aah60617 mus muscu
6	1404	42.8	644	2 Q700B9	Q7q0e9 anopheles g
7	1392.5	42.5	622	1 GL10_CAEEL	Q459q7 caenorhabdi
8	1360	41.5	666	1 GLT6_DROME	Q6wv16 drosophila
9	1360	41.5	666	2 AAQ56703	Aaq56703 drosophila
10	1341.5	40.9	599	2 Q7P2M5	Q7p2m5 anopheles g
11	1247	38.0	1003	2 Q7ODR0	Q7qd10 anopheles g
12	1186	36.2	659	1 GLT4_DROME	Q81a42 drosophila
13	1186	36.2	659	2 AAQ56701	Aaq56701 drosophila
14	1172.5	35.8	650	1 GLT5_DROME	Q8mcf3 m polypepti
15	1159	35.4	645	2 Q708T5	Q7q8t5 anopheles g
16	1144	34.9	518	2 Q7Q048	Q7q048 anopheles g
17	1125	34.3	559	1 GLT1_HUMAN	Q29142 h polypepti
18	1117	34.1	559	1 GLT1_PIG	Q29143 s polypepti
19	1116	34.0	559	1 GLT1_RAT	Q07537 b polypepti
20	1115	34.0	559	1 GLT1_BOVIN	Q08912 m polypepti
21	1115	34.0	559	1 GLT1_MOUSE	Q08912 m polypepti
22	1113	34.0	601	1 GLT7_CAEEL	Q6gm51 caenorhabdi
23	1112.5	33.9	617	2 AAQ56702	Aaq56702 drosophila
24	1111.5	33.9	617	2 AAQ56702	Aaq56702 drosophila
25	1109	33.8	556	1 GL13_MOUSE	Q8cfe3 r polypepti
26	1109	33.8	556	1 GL13_RAT	Q8ue29 r polypepti
27	1109	33.8	556	2 AAQ5749	Aaq5749 ratcus no
28	1106.5	33.7	630	1 GLT5_DROME	Q6wv17 drosophila
29	1106	33.7	556	1 GLT5_HUMAN	Q81ucv h polypepti
30	1087	33.2	559	2 O6P8A1	O6p8a1 xenopus lae
31	1087	33.2	559	2 AAH60419	Aah60419 xenopus l

32	1083	33.0	626	1 GLT5_CAEEL	Q95z11 caenorhabdi
33	1083	33.0	653	2 Q6N824	Q6n824 xenopus lae
34	1083	33.0	653	2 AAH70527	Aah70527 xenopus l
35	1053	32.1	653	2 Q6XK21	Q6xk21 xenopus tro
36	1053	32.1	653	2 AAH67317	Aah67317 xenopus t
37	1050	32.0	657	1 GLT7_HUMAN	Q86f52 homo sapien
38	1049	32.0	930	1 GLT5_MOUSE	Q8c102 m polypepti
39	1047	31.9	657	1 GLT7_MOUSE	Q80v40 mus musculu
40	1040	31.7	657	1 GLT7_RAT	Q9rc05 ratcus norv
41	1040	31.7	930	1 GLT5_RAT	Q88422 r polypepti
42	1037	31.6	606	2 Q7QEH0	Q7qeh0 anopheles g
43	1036	31.6	578	1 GLT4_HUMAN	Q8n448 h polypepti
44	1035.5	31.6	591	1 GLT4_DROME	Q8m448 drosophila
45	1035.5	31.6	591	2 AAQ56704	Aaq56704 drosophila

ALIGNMENTS

RESULT 1

ID	GL10_HUMAN	STANDARD	PRT	603 AA.
AC	Q868r1; Q6INS6; Q86Vp8; Q8IXJ2; Q8TEJ2; Q96IV2; Q9H8E1; Q9Y4M4;			
DT	01-OCT-2004 (Rel. 45, Created)			
DT	01-OCT-2004 (Rel. 45, Last sequence update)			
DE	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)			
DE	(protein-UDP acetylglucosaminyltransferase 10) (UDP-			
DE	GalNAc:polypeptide N-acetylglucosaminyltransferase 10) (Polypeptide			
DE	GalNAc transferase 10) (GalNAc-T10) (pp-GalTase 10).			
GN	Name=GAINT10;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM 1), ENZYME ACTIVITY, AND TISSUE			
RP	SPECIFICITY.			
RC	TISSUE=Colon cancer;			
RX	MEDLINE=22304871; PubMed=12417297;			
RA	Cheng L., Tachibana K., Zhang Y., Guo J.-M., Tachibana K.K.,			
RA	Kameyama A., Wang H., Hiruma T., Iwasaki H., Togayachi A., Kudo T.,			
RA	Narimatsu H.;			
RT	"characterization of a novel human UDP-GalNAc transferase, pp-GalNAc-			
RT	T10.";			
RL	FEBS Lett. 531:115-121(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RP	Bennett E.P.;			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM 4).			
RP	TISSUE=Uterus;			
RC	TISSUE=Uterus;			
RX	Medline=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;			
RA	Straussberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,			
RA	Klauser R.D., Collins F.S., Wagner L., Siemien C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Scaplenko M., Soares M.B., Bonaldi M.F., Casavant T.L., Schaefer T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Riba S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mulian S.J.,			
RA	Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Phley J., Helton B., Ketterman M., Madan A., Rodriguez S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RP [5]
 RP SEQUENCE OF 8-366 FROM N.A. (ISOFORM 3), AND SEQUENCE OF 117-603 FROM
 RP N.A. (ISOFORM 2).
 RC TISSUE=placenta, and Spleen;
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
 RA Shiratori A., Sudo H., Hosoki T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M.,
 RA Yamazaki M., Niimoriya K., Iehibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y.,
 RA Iehida S., Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hattori T.,
 RA Kusano Y., Kanehori K., Takahashi-Fujii A., Hara H., Tanase I.-O.,
 RA Nomura Y., Togashi K., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Iimose N., Musashino K., Yuuki F., Oshima A., Sasaki N., Aotsuka S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakabe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Komiyama K., Tashiro H., Tanigami A., Fujisawa T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Okumori Y.,
 RA Kawabata A., Hikiji T., Kobayashi N., Inagaki H., Ikemura Y., Okamoto S.,
 RA Ohtsuki R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togaishi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sigano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45(2004).
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein receptor. Has activity toward Muc5ac and Ez2 peptide
 CC substrates.
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=1;
 CC IsoId=Q86SR1-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q86SR1-2; Sequence=VSP_011209;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q86SR1-3; Sequence=VSP_011212; VSP_011213;
 CC Note=No experimental confirmation available;
 CC Name=4;
 CC IsoId=Q86SR1-4; Sequence=VSP_011207; VSP_011208; VSP_011214;
 CC Note=No experimental confirmation available;
 CC Name=5;
 CC IsoId=Q86SR1-5; Sequence=VSP_011210; VSP_011211;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed at high level in
 CC small intestine, and at intermediate levels in stomach, pancreas,
 CC ovary, thyroid gland and spleen. Weakly expressed in other

CC tissues.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called GT1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-Gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycoside specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -1- CAUTION: According to experiments made in rat, this enzyme is
 CC unable to transfer GalNAc onto serine or threonine residue on the
 CC protein receptor, but instead requires the prior addition of a
 CC GalNAc on a peptide before adding additional GalNAc moieties,
 CC thereby acting as a glycopeptide transferase.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL, AB078145; BAC56890.1; -
 CC EMBL, AJ505950; CAB44532.1; -
 CC EMBL, AL096739; CAB46378.1; -
 CC EMBL, BC007224; AAH07224.2; -
 CC EMBL, BC050333; AAH50333.1; -
 CC EMBL, BC072450; AAH72450.1; -
 CC EMBL, AK023782; BAB14676.1; -
 CC EMBL, AK074132; BAB84958.1; -
 CC PIR, T12552; T12552.
 CC HSSP, P26514; IKM.
 CC GeneW, HGNC:19873; GALNT10.
 CC MIM, 608043; -
 CC InterPro, IPR001173; Glyco trans 2.
 CC InterPro, IPR008997; RicinB like-
 CC InterPro, IPR000772; RicinB lectin.
 CC Pfam, PF00535; Glycoe_transf_2; 1.
 CC Pfam, PF00652; Ricin_B_lectin; 3.
 CC SMART, SM00458; Ricin, 1.
 CC PROSITE, PS02311; RICIN B LECTIN; 1.
 CC KW Alternative splicing; Calcium; Glycosyltransferase; Golgi stack;
 CC lectin; Manganese; Signal-anchor; Transferase; Transmembrane.
 CC FT DOMAIN 1 11
 CC FT TRANSMEM 12 31
 CC FT DOMAIN 32 603
 CC FT DOMAIN 144 253
 CC FT DOMAIN 311 373
 CC FT DOMAIN 458 590
 CC FT DISULFID 471 488
 CC FT DISULFID 523 538
 CC FT DISULFID 563 578
 CC FT CARBOHYD 124 146
 CC FT CARBOHYD 146 166
 CC FT CARBOHYD 593 593
 CC FT CARSPHYD 1 329
 CC FT VARSPLIC 330 352
 CC FT VARSPLIC 190 251
 CC FT VARSPLIC 190 202
 CC FT VARSPLIC 203 603
 CC FT VARSPLIC 354 366
 CC -----
 CC /FTId=VSP_011208.
 CC Missing (in isoform 2).
 CC /FTId=VSP_011209.
 CC EHLKPLBDYML -> DLPASTSPFVC (in
 CC isoform 5).
 CC /FTId=VSP_011210.
 CC Missing (in isoform 5).
 CC /FTId=VSP_011211.
 CC WMGGDPGLIEGGEYVSRK -> MLWRDGELEARTS
 CC SGLFLIWM (in isoform 4).
 CC /FTId=VSP_011208.
 CC Missing (in isoform 2).
 CC /FTId=VSP_011209.
 CC EHLKPLBDYML -> DLPASTSPFVC (in
 CC isoform 5).
 CC /FTId=VSP_011210.
 CC Missing (in isoform 5).
 CC /FTId=VSP_011211.
 CC WMGGDPGLIEGGEYVSRK -> MLWRDGELEARTS
 CC SGLFLIWM (in isoform 4).

Query Match 96.3%; Score 3157; DB 1; Length 603;
 Best Local Similarity 96.0%; Pred. No. 2e-247;
 Matches 579; Conservative 7; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERROPDGTGGGAAPAAAGGSHSRK 60
 DB 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERROPDGTGGGAAPAAAGGSHSRK 60
 QY 61 KTFPLGDGQKLDKMDHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 DB 61 KTFPLGAEQRLKDMHNKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 QY 121 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 180
 DB 121 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 180
 QY 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 240
 DB 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 240
 QY 241 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 300
 DB 241 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 300
 QY 301 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 DB 361 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 QY 421 AVQKRLSSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 480
 DB 421 AVQKRLSSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 480
 QY 481 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 540
 DB 481 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 540
 QY 541 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 600
 DB 541 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 3
 BAD21405 PRELIMINARY; PRT; 634 AA.
 ID BAD21405.
 AC BAD21405.
 DT 01-JUN-2004 (T-EMBLrel. 27, Created)
 DT 01-JUN-2004 (T-EMBLrel. 27, Last sequence update)
 DE MFLJ00205 protein (Fragment).
 GN MFLJ00205.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBITaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Okazaki N., Kitano R., Ohara R., Inamoto S., Koseki H., Hirakata S.,
 RA Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.,
 RT "Prediction of the Coding Sequences of Mouse Homologues of FlJ Gene:
 RT The Complete Nucleotide Sequences of 110 Mouse FlJ-homologous cDNAs
 RT Identified by Screening of Terminal sequences of cDNA Clones Randomly
 RT Sampled from Size-Fractionated Libraries";
 RT Submitted (FE8-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK131155; BAD21405.1; -.
 FT NON_TER 1

SQ SEQUENCE 634 AA; 72315 MW; C79B82D4D0052C81 CRC64;
 Query Match 95.7%; Score 3138; DB 2; Length 634;
 Best Local Similarity 95.5%; Pred. No. 7.6e-246;
 Matches 576; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKEKRLLOAVLVAALVLLPNVGLMALYERROPDGTGGGAAPAAAGGSHSRK 60
 DB 32 MRKEKRLLOAVLVAALVLLPNVGLMALYERROPDGTGGGAAPAAAGGSHSRK 91
 QY 61 KTFPLGDGQKLDKMDHKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 120
 DB 61 KTFPLGAEQRLKDMHNKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 QY 92 KTFPLGAEQRLKDMHNKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 DB 92 KTFPLGAEQRLKDMHNKEAIRDAQVNGEGQRPVPMTDABRVDAVRENGFNIVSDK 151
 QY 121 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 180
 DB 121 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 211
 QY 152 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 211
 DB 152 ISLNRLPDIRHPNCSKRYLETLPNTSIIIPHNHNGSSLLRTVSVLNRSPELVAEI 211
 QY 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 240
 DB 181 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 271
 QY 212 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 271
 DB 212 VLVDDSDREHLKKPLEDYMALPPSVRIILRTKREGILRTMLGASAAATGDTITPLDSHC 271
 QY 241 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 300
 DB 241 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 331
 QY 272 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 331
 DB 272 EAVNMLPPLLDRIANRRTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEYTKKIP 331
 QY 301 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 391
 QY 332 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 391
 DB 332 PELQKADPSDPFESPVMAGGLFVDRKFWELGSDYDGLGEGEYEISFKVMCGGRM 391
 QY 361 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 420
 DB 361 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 451
 QY 392 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 451
 DB 392 EDIPCSRVGHIYRKVYPPVPAVSLARNLKRVAEVMDEYAEYIQRREYRHLSAGDV 451
 QY 421 AVQKRLSSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 480
 DB 421 AVQKRLSSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 511
 QY 452 VQKRLRVSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 511
 DB 452 VQKRLRVSLNCKSPFKMFTKIAMDLPKFYPPVEPPAAAGSIRNVTGLCADTKGALG 511
 QY 481 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 540
 DB 481 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 571
 QY 512 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 571
 DB 512 SPRLLETCTIRGGEAAMNSQVFTFTWRREDIRPGDPQHTKKCFDVAHSITSPTLYDCHS 571
 QY 541 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 600
 DB 541 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 631
 QY 572 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 631
 DB 572 MKGNQIMKYRKDXTLVHPVSGSCMDCSDHRIFMNTCPSSLTQWLFHTNSTYLENF 631
 QY 601 NRN 603
 DB 601 NRN 603
 QY 632 NRN 634
 DB 632 NRN 634

RESULT 4
 GL10_MOUSE STANDARD; PRT; 603 AA.
 ID GL10_MOUSE
 AC Q6P9S7; Q6KAQ2; Q8B208; Q91YU6;
 DT 01-OCT-2004 (Rel. 45, Created)
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DE 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Polypeptide N-acetylglucosaminyltransferase 10 (EC 2.4.1.41)
 DE (Protein-UDP acetylglucosaminyltransferase 10 (UDP-
 DE GAlNAc:Polypeptide N-acetylglucosaminyltransferase 10) (Polypeptide
 DE GAlNAc transferase 10) (GAlNAc-T10) (pp-GAlNAc 10).
 GN Name=GAlNAcT10;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBITaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Okazaki N., Kitano R., Ohara R., Inamoto S., Koseki H., Hirakata S.,
 RA Saga Y., Kitamura H., Nakagawa T., Nagase T., Ohara O., Koga H.,
 RT "Prediction of the coding sequences of mouse homologues of FlJ genes:
 RT the complete nucleotide sequences of 110 mouse FlJ-homologous cDNAs

RT identified by screening of terminal sequences of cDNA clones randomly
 RT sampled from size-fractionated libraries.";
 RL DNA Res. 11:167-180(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6, and FVB/N; TISSUE=Brain, and Breast tumor;
 RX MEDLINE=22358257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueffing T.B., Tothiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,
 RA Bosak S.A., McEwan P.D., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richardson S., Wooley K.C., Hale S., Garcia A.M., Gay L.U., Hultky S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schultz J., Myers R.W.,
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schneroch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 76-603 FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Colon;
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamana H., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schombach C., Gojohori T.,
 RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schirral L.M., Kanadin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bratt D., Brusic V., Chochia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Glasi C., Godzik A., Gough J.,
 RA Grønborg S., Guenlinclon S., Hirokawa N., Jackson I.U., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King P.A.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pettes G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius U.U., Qi D., Ramchandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shimada M.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
 RA Varrault R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazune N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Atzawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Ysuntshi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21264503; PubMed=11278534; DOI=10.1074/jbc.M009638200;
 RA Ten Hagen K.G., Bedi G.S., Tetsart D., Kingsley P.D., Hagen F.K.,
 RA Balys M.M., Beres T.M., Degand P., Tabak L.A.;
 RT "Cloning and characterization of a ninth member of the UDP-
 RT GalNAc:polypeptide N-acetylglucosaminyltransferase family,"
 RL J. Biol. Chem. 276:17395-17404(2001).
 CC -1- FUNCTION: Catalyzes the initial reaction in O-linked
 CC oligosaccharide biosynthesis, the transfer of an N-acetyl-D-
 CC galactosamine residue to a serine or threonine residue on the
 CC protein receptor. Has activity toward Muc5Ac and EA2 peptide

CC substrates (By similarity).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 CC UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed at higher level than GALNT9. In the
 CC developing hindbrain region of E14.5 embryos it accumulates in the
 CC rapidly dividing, undifferentiated ventricular zone adjacent to
 CC the pons. It also accumulates in the regions immediately rostral
 CC and caudal to the dorsal rhombic lips differentiating into the
 CC cerebellum. Not expressed in the developing choroid plexus.
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 CC region: the N-terminal domain (domain A, also called G1 motif),
 CC which is probably involved in manganese coordination and substrate
 CC binding and the C-terminal domain (domain B, also called
 CC Gal/GalNAc-T motif), which is probably involved in catalytic
 CC reaction and UDP-Gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 CC contributes to the glycopeptide specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 CC subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -1- CAUTION: According to experiments made in rat, this enzyme is
 CC unable to transfer GalNAc onto serine or threonine residue on the
 CC protein receptor, but instead requires the prior addition of a
 CC GalNAc on a peptide before adding additional GalNAc moieties,
 CC thereby acting as a glycopeptide transferase.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AK131155; BA021405.1; ALT INIT.
 CC EMBL, BC016585; AAH16585.1; ALT INIT.
 CC EMBL, BC060617; AAH0617.1; -
 CC EMBL, AK033515; BA028334.1; -
 CC HSSP; P26514; 1KNM.
 CC MD5: MGI:1890480; Galnt10.
 CC GO: GO:0004653; F:polypeptide N-acetylglucosaminyltransferase. . . ; IDA.
 CC GO: GO:0006493; P:O-linked glycosylation; IDA.
 CC InterPro: IPR001173; Glyco trans 2.
 CC InterPro: IPR008997; RicinB like.
 CC InterPro: IPR000772; Ricin B lectin.
 CC Pfam: PF00535; Glycos transf_2; 1.
 CC Pfam: PF00652; Ricin B lectin; 2.
 CC SMART; SM00458; RICIN; 1.
 CC PROSITE; PSS0231; RICIN B LECTIN; 1.
 CC KEGG; Glycosyltransferase; Golgi stack; Lectin; Manganese;
 CC KEGG; Signal-anchor; Transferase; Transmembrane.
 CC FT DOMAIN 1 11 Cytoplasmic (Potential).
 CC FT TRANSMEM 12 31 Signal-anchor for type II membrane
 CC FT protein (Potential).
 CC FT DOMAIN 32 603 Luminal (Potential).
 CC FT DOMAIN 311 373 Catalytic subdomain A.
 CC FT DOMAIN 458 590 Ricin B-type lectin.
 CC FT DISULFID 471 488 By similarity.
 CC FT DISULFID 523 538 By similarity.
 CC FT CARBOHYD 124 124 N-linked (GlcNAc. . .) (Potential).
 CC FT CARBOHYD 146 146 N-linked (GlcNAc. . .) (Potential).
 CC FT CARBOHYD 593 593 N-linked (GlcNAc. . .) (Potential).
 CC FT CONFLICT 233 233 V -> I (in Ref. 1).
 CC SQ SEQUENCE 603 AA; 6916 MW; FF55FBA7E1DD7544 CRC64;
 CC Query Match 95.7%; Score 3137; DB 1; Length 603;
 CC Best Local Similarity 95.4%; Pred. No. 8.6e-246;

Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKREKRLQAVLVAALVLPVNGLMALYREROPDGTGCGSAAVAPAAAGGSHSROK 60
 DB 1 MRKREKRLQAVLVAALVLPVNGLMALYREROPDGTGCGSAAVAPAAAGGSHSROK 60
 QY 61 KTFEFLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 DB 61 KTFEFLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 QY 121 ISLNRLPDIRHNCNCKSLYLETLPNTSIIIPFHNEGSSLLRTYHVSVLRSPPELVARI 180
 DB 121 ISLNRLPDIRHNCNCKSLYLETLPNTSIIIPFHNEGSSLLRTYHVSVLRSPPELVARI 180
 QY 181 VLVDPSDREHKKPLEDYMALPSSVRIARTKREGILRTMGLGASVATGDTVTFDLSHC 240
 DB 181 VLVDPSDREHKKPLEDYMALPSSVRIARTKREGILRTMGLGASVATGDTVTFDLSHC 240
 QY 241 EAVNWMPLPDLRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMETYYKRIPIR 300
 DB 241 EAVNWMPLPDLRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMETYYKRIPIR 300
 QY 301 PELQKADPSDPSPSPVMAAGLFAVDKRMFELGYPGLEINGEGYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPSPSPVMAAGLFAVDKRMFELGYPGLEINGEGYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGV 420
 DB 361 EDIPCSRVGHIYRKVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGV 420
 QY 421 AVQKRLRSSLNCKSPKFMFTKIAMDLPKFPYPVEPPAAAGSIRVGTGLCADTKHGALG 480
 DB 421 AVQKRLRSSLNCKSPKFMFTKIAMDLPKFPYPVEPPAAAGSIRVGTGLCADTKHGALG 480
 QY 481 SPRLRLEGCVRGGEAAMNNNOVFTFWREDIRPGDQHTKCFPDALSHSPPTLYDCHS 540
 DB 481 SPRLRLEGCVRGGEAAMNNNOVFTFWREDIRPGDQHTKCFPDALSHSPPTLYDCHS 540
 QY 541 MKNQMLMKRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTOOMLFPHNTSYLVEKF 600
 DB 541 MKNQMLMKRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTOOMLFPHNTSYLVEKF 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 5
 AAH60617 PRELIMINARY; PRT; 603 AA.

AC AAH60617
 DT 12-MAY-2004 (TEMBLrel. 27, Created)
 DT 12-MAY-2004 (TEMBLrel. 27, Last sequence update)
 DE UDP-N-acetyl-alpha-D-galactosamine:polypeptide
 DE N-acetyl-galactosaminyltransferase 10.
 GN GALNT10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCB1_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haefl F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richard S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Vailion D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting J., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Smyth J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalska U., Smallov D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RU Submitted (Oct-2003) to the EMBL/GenBank/DBJ database.
 DR EMBL; BC060617; AAH60617.1; -.
 KW transferase.
 SQ SEQUENCE 603 AA; 6916 MW; FFS5FBA7E1DD7544 CRC64;

Query Match 95.7%; Score 3137; DB 2; Length 603;
 Best Local Similarity 95.4%; Pred. No. 8.6e-246;
 Matches 575; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRKREKRLQAVLVAALVLPVNGLMALYREROPDGTGCGSAAVAPAAAGGSHSROK 60
 DB 1 MRKREKRLQAVLVAALVLPVNGLMALYREROPDGTGCGSAAVAPAAAGGSHSROK 60
 QY 61 KTFEFLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 DB 61 KTFEFLDGGQKLDKMDHKEAIRDAQVNGEGQRPYPMTDAERVDQAYRENGENIYVSDK 120
 QY 121 ISLNRLPDIRHNCNCKSLYLETLPNTSIIIPFHNEGSSLLRTYHVSVLRSPPELVARI 180
 DB 121 ISLNRLPDIRHNCNCKSLYLETLPNTSIIIPFHNEGSSLLRTYHVSVLRSPPELVARI 180
 QY 181 VLVDPSDREHKKPLEDYMALPSSVRIARTKREGILRTMGLGASVATGDTVTFDLSHC 240
 DB 181 VLVDPSDREHKKPLEDYMALPSSVRIARTKREGILRTMGLGASVATGDTVTFDLSHC 240
 QY 241 EAVNWMPLPDLRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMETYYKRIPIR 300
 DB 241 EAVNWMPLPDLRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWMETYYKRIPIR 300
 QY 301 PELQKADPSDPSPSPVMAAGLFAVDKRMFELGYPGLEINGEGYEISFKVMCGGRM 360
 DB 301 PELQKADPSDPSPSPVMAAGLFAVDKRMFELGYPGLEINGEGYEISFKVMCGGRM 360
 QY 361 EDIPCSRVGHIYRKVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGV 420
 DB 361 EDIPCSRVGHIYRKVPYKVPAGVSLARNLKRVAEVMDEYAEIYQRRPEYRHLISAGV 420
 QY 421 AVQKRLRSSLNCKSPKFMFTKIAMDLPKFPYPVEPPAAAGSIRVGTGLCADTKHGALG 480
 DB 421 AVQKRLRSSLNCKSPKFMFTKIAMDLPKFPYPVEPPAAAGSIRVGTGLCADTKHGALG 480
 QY 481 SPRLRLEGCVRGGEAAMNNNOVFTFWREDIRPGDQHTKCFPDALSHSPPTLYDCHS 540
 DB 481 SPRLRLEGCVRGGEAAMNNNOVFTFWREDIRPGDQHTKCFPDALSHSPPTLYDCHS 540
 QY 541 MKNQMLMKRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTOOMLFPHNTSYLVEKF 600
 DB 541 MKNQMLMKRKDKTLVHPVSGSCMDCESDHRIFMNTCNPSLTOOMLFPHNTSYLVEKF 600
 QY 601 NRN 603
 DB 601 NRN 603

RESULT 6
 Q700E9 PRELIMINARY; PRT; 644 AA.

AC Q70E9; 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE AGCP9480 (Fragment).
GN Name=agcG54007; ORFNames=ENSANGG0000009226;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygotia;
CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/genbank/DBJ databases.
CC -1 CAUTION: The sequence shown here is derived from an
EMBL/genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008986; EAA00190.1; -
DR InterPro; IPR001173; Glyco_trans_2.
DR InterPro; IPR007772; Ricin_B_lectin.
DR Pfam; PF00535; Glycos_transf_2.1.
DR Pfam; PF00652; Ricin_B_lectin; 3.
DR PROSITE; PS0231; RICIN_B_LECTIN; 1.
FT NON TER 1
SQ SEQUENCE 644 AA; 74214 MW; 0B7AB007BC228501 CRC64;

Query Match 42.8%; Score 1404; DB 2; Length 644;
Best Local Similarity 45.6%; Pred. No. 5,6e-105;
Matches 283; Conservative 104; Mismatches 195; Indels 38; Gaps 16;

QY 1 MRREKRLQLVAVLVLAIVLPNVGLMALIRERPPDGTGGSGAANVAAPAGCGSHSKQ 60
DB MRRNRVTLIKITLYFSGALLFI-----TLIRSPSDAKNTLDGVFPPEQMCKESHAR 89
QY 61 KTFPLGD-----GOLKDMHDKEARDAORVGENEGRPFYM--TDAREVDQAARENGF 113
DB EGSEFFNNNSKVHVQRIDMHNVDLIHEAKSGISEHGAGOLDKSEHKMDKPKXNGF 149
QY 114 NIYVSDKISLRSFLDIRHPNCSKRYLETPNTSIIIPFHNEGSSILRTVHSVLNRP 173
DB 150 NAVLSDKISLRSFLDIRHRGRKKQIYSELPTYSVVVPFYNEHSTLLRTASSVLNRP 209
QY 174 PELVAELIVDDPSDRHLKPLEDYMA-LFPSVILRTKKREGILIRFMIGASVATGDV 232
DB 210 PELIAELLIVDDCSKEFKOQLDEVYTENMPKVAVRLPERSGLITARIAGAATAADV 269
QY 233 ITPLDSHCNAVNMVPLLDIRIARNRKTIYCPMIDVIDHDPRYTOAGDMRGAFEDWM 292
DB 270 LIFLDSHTEAVNMVPLLEPIAEYRYCVCCEFDIVIDWDFFEYQAQ--DEGARGFADKF 328
QY 293 YYKRIP-IPELOKADPSDPESPVMAAGLFVADRKMFEWLEGVDPGLEIWGEQYEISF 351
DB 329 FYKRLPLPRDQ-NPTPEPESPIMAGGLPAISAKFWETIGYBEGDINGEBYEISF 386
QY 352 KVMMCGRMEDIPCSRVGHIYRKVYPYKVPAGVS-LARNLRVAEVMDEYAELYQR-R 409
DB 387 KIWGCGGMVYDAPCSRVGHIYRGVAPFGNPFRKDELITNYKYKVAEVMDEYEXLYMADR 446
QY 410 PEYRLSLAGDVAIVOKKLSSLNCSFKPMTKIANDLKFPYRVVPPRPAAMGEIRNV-GT 468
DB 447 KYEYETDVGDISRQLAIRKQLCKEPKMFMTQVAVDLIEKPRIIPPFAAGALQSIVANA 506
QY 469 GLCADTKKGALGPSPLTEGCYVGRGEAAMNNMVQVTFPMREDIRBGDPOHTKKF---CFD 525
DB 507 ALCVDTLNAGEKQTIGLYSCAEDKKQPQN-QFOLSMHRDLR-----IKBELCMD 557
QY 526 AISHSP---VTLYDCHSNKNQMLKYRKD-KTLYHPVSGSCMDCSESDHRIFNNTCNPS 581
DB 558 -VSESVPAKILLHYCHGGQGQNLWRYPEBTOMLKQGGKNNRCLDWNPNNRBEVFVNPCPT 616
QY 582 SLTGOWMEFHNTSYLLEKEN 601

Dn		617 NPROKMRGFINATSLAQM^N 636
	RESULT 7	
AC	GL10_CAEEL	STANDARD; PRT; 622 AA.
AD	045947;	
DT	01-OCT-2004 (Rel. 45, Created)	
DT	01-OCT-2004 (Rel. 45, Last sequence update)	
DT	01-OCT-2004 (Rel. 45, Last annotation update)	
DE	Putative polypeptide N-acetylglucosaminyltransferase 10	
DE	(EC 2.4.1.41) [Protein-UDP acetylglucosaminyltransferase 10] (UDP-	
DE	GalNAc:polypeptide N-acetylglucosaminyltransferase 10) [pp-GaNtase	
DN	10].	
GN	Name=gly-10; ORFName=Y45F10D.3;	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;	
OC	Rhabditiidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RA	McMurray A.A.;	
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.	
CC	- FUNCTION: May catalyze the initial reaction in O-linked	
CC	oligosaccharide biosynthesis, the transfer of an N-acetyl-D-	
CC	galactosamine residue to a serine or threonine residue on the	
CC	protein receptor (By similarity).	
CC	- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =	
CC	UDP + N-acetyl-D-galactosaminyl-polypeptide.	
CC	- COFACTOR: Manganese and calcium (By similarity).	
CC	- PATHWAY: Glycosylation.	
CC	- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By	
CC	similarity).	
CC	- DOMAIN: There are two conserved domains in the glycosyltransferase	
CC	region: the N-terminal domain (domain A, also called Grl motif),	
CC	which is probably involved in manganese coordination and substrate	
CC	binding and the C-terminal domain (domain B, also called	
CC	Gal/GalNAc-T motif), which is probably involved in catalytic	
CC	reaction and UDP-Gal binding (By similarity).	
CC	- DOMAIN: The ricin B-type lectin domain binds to GalNAc and	
CC	contributes to the glycopeptide specificity (By similarity).	
CC	- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T	
CC	subfamily.	
CC	- SIMILARITY: Contains 1 ricin B-type lectin domain.	

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/isb-sib.ch).	
CC	or send an email to license@isb-sib.ch).	

DR	EMBL, AL021492; CAAL6378.1; ALT_INIT.	
DR	PIR; T26930; T26930.	
DR	WormPep; Y45F10D.3; CEI6642.	
DR	InterPro; IPR001173; GIACO_trans_2.	
DR	InterPro; IPR008997; RicinB_like.	
DR	InterPro; IPR000772; Ricin_B_lectin.	
DR	Pfam; PF00535; Glycos_transf_2; 1.	
DR	Pfam; PF00652; Ricin_B_lectin; 3.	
DR	SMART; SM00458; RICIN; 1.	
DR	PROSITE; PS50231; RICIN_B_LECTIN; 1.	
KW	Calcium; Glycosyltransferase; Golgi stack; Hypothetical protein;	
KW	Lectin; Manganese; Signal-anchor; Transferase; Transmembrane.	
FT	DOMAIN	1
FT	TRANSMEM	13 .. 32
FT		Cytoplasmic (Potential).
FT		Signal-anchor for type II membrane
FT	DOMAIN	33 .. 622
FT		proteins (Potential).
FT	DOMAIN	156 .. 268
FT		luminal (Potential).
FT	DOMAIN	324 .. 386
FT		Catalytic subdomain A.
FT	DOMAIN	519 .. 622
FT		Ricin B-type lectin.

```

FT DISULFID 532 549 By similarity.
FT DISULFID 575 591 By similarity.
FT CARBOHYD 136 136 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 170 170 N-linked (GlcNAc...) (Potential).
SQ SEQUENCE 622 AA; 71811 MW; F8A3A21159C7B452 CRC64;

Query Match 42.5%; Score 1392.5; DB 1; Length 622;
Best Local Similarity 48.8%; Pred. No. 4.5e-104;
Matches 279; Conservative 87; Mismatches 175; Indels 31; Gaps 14;

QY PAAGQSHSRQKTFPLGQOKLQWMDKAIKRDQAVNGQGRPYPTMDERDQ--106
DB PPALDDEALDPFEKRGH-EKIK-WEDEAAYEKERKREBQGWKRVKLPEDKEVEKA 116
QY 107 --AYRNGENIYVSDKISLNRSLPDIRHPCNSKRYLETLPNTSIIPFNHGMSSILRT 164
DB 117 LSIYKANGNYAISDMISLNRSLKDIRHPCNSKRYLETLPNTSIIPFNHGMSSILRT 176
QY 165 VHSVLNRSPPELVAEIVLVDPSDRHLLKKPLEDY--ALPFSVILRTKKRGLIRTR 221
DB 177 VSVINRSPPELKEIILVDPSFSEKDALRQLEDFLKKNIIDHIVKVLRTKKRGLIRGR 236
QY 222 MGLASVATGDTVITFLSHSCANNWMLPILDRILARRKRTIVCMIDVIDHDDPRYETQAG 281
DB 237 QLGADATGEIILFLAHSEANWMLPILDRILAEYRTVVCPEVVIDDEYEVBPQ-D 295
QY 282 DAMRGAFDWMYKRIPIPELQKADPSDFESPVWAGLFAVDKRMFELAGYDGLRI 341
DB 296 EGRGSGFDMAFNKRLPLTKK-DRESPTKFPNSPVHAGYFALSAKMFELGIDGLDI 354
QY 342 WGGQYQYISFKWMCGRMEDIPCSRHYIYR-KYVPYK-VPAVSLARLTKFAEVMND 399
DB 355 WGGQYQYISFKWMCGRMEDIPCSRHYIYR-KYVPYK-VPAVSLARLTKFAEVMND 414
QY 400 EYAEYIYQRREPRHLSAGVANOKLRSLNCKSKRMFTKTLAMLPKRYPEVEPPAA 459
DB 415 DYKRETIYKRRPGVGNADAGLKLKMKIGIRELQCKSPDWFKETLAFQDKRYPAVEPKA 474
QY 460 WGEIRVGTGLCADTKHGLSGPLREGCVR---GRGEAANNMNOVFRTMEDIRPQG 515
DB 475 EGEIRVGTGLCADTKHGLSGPLREGCVR---GRGEAANNMNOVFRTMEDIRPQG 515
QY 516 POKTKKCFDALSHT--SPVTLVDCHSMKGNQMLKTR-KDKTLVHPVSGSCMDSCSSDHR 572
DB 527 --KGRKICFDCSTSVKAPVILFDCHSMKGNQMLFKTRVAKQIYHPIISGGLTADENK 584
QY 573 -IFMNTCNPSLTQQLFHTNSTVLEKFRN 603
DB 585 FLHMKKCDSSDLQKMAQTVNDELLETRQAN 616

RESULT 8
GLT6_DROME STANDARD; PRT; 666 AA.
AC Q6WV16; Q95R40; Q9VZK5;
AD 01-OCT-2004 (Rel. 45, Created)
DT 01-OCT-2004 (Rel. 45, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DE N-acetyl-galactosaminyltransferase 6 (EC 2.4.1.-) (Protein-UDP
DE acetyl-galactosaminyltransferase 6) (UDP-GalNAc:polypeptide N-
DE acetyl-galactosaminyltransferase 6) (pp-GANTase 6).
GN Namespant6; ORFNames-CG2103;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
RP DEVELOPMENTAL STAGE.
RC STRAIN=Canton-S; TISSUE=Embryo;
RX MEDLINE=22841110; PubMed=12829714; DOI=10.1074/jbc.M303836200;
RA Ten Hagen K.G., Tran D.T., Gerken T.A., Stein D.S., Zhang Z.;

```

```

RT RT "Functional characterization and expression analysis of members of the
RT UDP-GalNAc:polypeptide N-acetyl-galactosaminyltransferase family from
RT Drosophila melanogaster."
RT J. Biol. Chem. 278:35039-35048(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Cealinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.B., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Abgrayani A., An H.-J., Andrews-Pfeankoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pabloes B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Galburt W.M., Glasser K.,
RA Glodok A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kension J.A., Ketchum K.A.,
RA Klamel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li U.H., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclob J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saundere R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Sprelding A.C., Stapleton M., Strong R., Sun E.,
RA Stykars R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Krommiller B., Paclob J.M., Park S., Wan K.H.,
RA Rubin G.M., Cealinker S.E.;
RT "A Drosophila full-length cDNA resource."
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- FUNCTION: Glycopeptide transferase involved in O-linked
CC oligosaccharide biosynthesis, which catalyzes the transfer of an
CC N-acetyl-D-galactosamine residue to an already glycosylated
CC peptide. In contrast to other proteins of the family, it does not
CC act as a peptide transferase that transfers GalNAc onto serine or
CC threonine residue on the protein receptor, but instead requires
CC the prior addition of a GalNAc on a peptide before adding
CC additional GalNAc moieties. Some peptide transferase activity is
CC however not excluded, considering that its appropriate peptide
CC substrate may remain unidentified. Prefers the diglycosylated
CC Nucleoside-3/13 as substrate.
CC -1- COFACTOR: Manganese and calcium (By similarity).
CC -1- PATHWAY: Glycosylation.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
CC similarity).
CC -1- TISSUE SPECIFICITY: In embryo, it is specifically expressed in
CC the salivary glands from stage 12, becoming stronger at stage 13.
CC Not expressed in other tissues. Further expressed during

```


QY 243 NNNMLPDLRIARNKRTIVCPMIDVIDHDDFFRYETOAGDAMGAPDMWYKRIPIPE 302
 DB 301 NNNMLPDLRIARNKRTIVCPMIDVIDHDDFFRYETOAGDAMGAPDMWYKRIPIPE 359
 QY 303 LQADSDSDPESVMAAGLFAVDRKMFWEIGYDPLGLETIGSGOYELSPFVMMCGGEMD 362
 DB 360 DLK-HPADPESKSDIMAGGLFAISREFPEWELGGYDEGLDIDGGEOYELSPFVMMCGGEMD 418
 QY 363 IPCSRGHIYR-----KYVYKVPAGVSLARNLKRVAEVMWDEYAEIYORRE-FRHL 416
 DB 419 APCSRIIGHYGRNHOPSBRK---GDYLNKTKRYAEVMMDEYKNTLYLSHGGLYESVD 475
 QY 417 AGDVAQKLRSSINCKSPFKMFKMTKIAMDLPKFYPVPEPPAAAGELRYNGT-GLCADT- 474
 DB 476 PGDLTGQKAIKRTLNCSFPCMFMEVAFDLMTKYTPVPDPSYAMGALQNGNQLCDTL 535
 QY 475 ---KHGALG-----SPRLBEGCYRGGEAAMNMVFTFTWREDIRPGDPQHTKFC 522
 DB 536 GRKHNMKMYACADNIKTFQR-----TQPMELSWKRDLE---LRKKE 576
 QY 523 CPDA--ISHTSPVTLVDCSHMKNQLMKY-RKDKTLVHPVS-GCMDCSSDHRIFMNTC 578
 DB 577 CLDVQIWDANAPVWMDCHSGOGNQYIYRYRHKOLKRGTRKCLLELPSEGVANKC 636
 QY 579 NPSSLTQOMLFEHTNSTVLEKENRN 603
 DB 637 DTDNRFQNNFGSPNKALDNYGSD 661

RESULT 10

Q7PZMS PRELIMINARY; PRT; 599 AA.
 ID Q7PZMS; AC Q7PZMS;
 DT 01-MAR-2004 (TEMBLrel. 26, Created)
 DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE AGCP9340 (Fragment).
 GN Name=agCG54449; ORFNames=ENSANG0000012813;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota,
 OC Neoptera, Endopterygota, Diptera, Nematocera, Culicoidae, Anopheles.
 NC NCB1_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: AAB01008986; EAA00339.1; -;
 DR InterPro: IPR001173; Glyco_transf. 2.
 DR InterPro: IPR000772; Ricin_B_lectin.
 DR Pfam: PF00535; GlycoB_transf. 2; 1.
 DR Pfam: PF00652; Ricin_B_lectin; 3.
 DR PROSITE: PS50231; RICIN_B_LECTIN; 1.
 FT NON_TER 1
 SQ SEQUENCE 599 AA; 69224 MW; DA646C182B143028 CRC64;

Query Match 40.9%; Score 1341.5; DB 2; Length 599;
 Best Local Similarity 45.8%; Pred. No. 6e-100;
 Matches 257; Conservative 104; Mismatches 175; Indels 25; Gaps 13;
 QY 52 GGGSHSQKTFPLGSGQKXKMDHDKAIRDAROVNGSGRPYMTDAERY---DQAY 108
 DB 46 GGEFYAPRNIV---AGEKI-DWNNYELIEESRGTGSGHGHPYKLSQDIALANKLP 100
 QY 109 RENGFIYVSDKISLNSLPLDIRHPNCSKRYLETLPNTSIIIPFHEGWSLLRTVHSV 168
 DB 101 KENGYSAYVSDMALNRSVPDIRHPSCKMKEYLKELPVSVIIIFVHEHMSALLRTVSV 160
 QY 169 LNRSPPELVAEIVLVDDFSDEHKLKPLEDYM--ALFPSVRIILTKRKEGLIRTMIGAS 226

DB 161 LNRSPPALKEIILVNDHSTKPELWTPLEFVSEELAPRYRLVDDPEBSGLIVARAGAR 220
 QY 227 VATGDIVTEFHCSEANVMMLPDLRIARNKRTIVCPMIDVIDHDDFFRYETOAGDAMG 286
 DB 221 EARGDVILVDBHTENTWMLPEPLERLIEDYRTVCPTVIDIAHDTFQYRQ-DEGRG 279
 QY 287 APDMWYKRIIP-IPPELOKADSDPESVMAAGLFAVDRKMFWEIGYDPLGLETIGSGE 345
 DB 280 AFDWKYRYRLPLLPDLD--DPTKPFNSPVMAAGLFAISAKFWEIGYDGLDWGE 337
 QY 346 QYELSPKVMCGGRMEDTCSRVGHYKRYVYKVPAGIS-LARNLKRVAEVMWDEYAEY 404
 DB 338 QYELSPKVMCGGRMEDTCSRVGHYKRYVYKVPAGIS-LARNLKRVAEVMWDEYAEY 397
 QY 405 IYORREYVHLSAGDVAQKLRSSINCKSPFKMFKMTKIAMDLPKFYPVPEPPAAAGELR 464
 DB 398 LYERNQPAKTPBGDLSAORBLERLQCFPFKMLEVAVPDLVRYPPDPQPFASGRVQ 457
 QY 465 NVGT-GLCADTKHAGLSEPLREGCYRGGEAAMNMVFTFTWREDIRPGDPQHTKFC 523
 DB 458 SVANPRLCDSLNHQKEDIGLYACAFNKTHP--QNNQFFTLSYHDIR---VRSNDK 511
 QY 524 PDAISHTSPVTLVDCSHMKNQLMKY-RKDKTLVHPVS--GCMDCSSDHRIFMNTC 580
 DB 512 LPAALNDRIVLFSCHESQNMWRYDYESKMIHGKHGRGMEADLHTNLFPVASCOK 571
 QY 581 SSLTQOMLFEHTNSTVLEKENRN 601
 DB 572 DKASQKNNMGYNFVHLQWMD 592

RESULT 11

Q7ODRO PRELIMINARY; PRT; 1003 AA.
 ID Q7ODRO; AC Q7ODRO;
 DT 01-MAR-2004 (TEMBLrel. 26, Created)
 DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
 DE AGCP10665 (Fragment).
 GN Name=agCG47419; ORFNames=ENSANG0000013497;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota,
 OC Neoptera, Endopterygota, Diptera, Nematocera, Culicoidae, Anopheles.
 NC NCB1_TaxID=180454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: AAB01008849; EAA07231.1; -;
 DR InterPro: IPR001173; Glyco_transf. 2.
 DR InterPro: IPR000772; Ricin_B_lectin.
 DR Pfam: PF00535; GlycoB_transf. 2; 2.
 DR Pfam: PF00652; Ricin_B_lectin; 4.
 DR PROSITE: PS50231; RICIN_B_LECTIN; 2.
 FT NON_TER 1
 SQ SEQUENCE 1003 AA; 115923 MW; 753EA50F567A4E13 CRC64;

Query Match 38.0%; Score 1247; DB 2; Length 1003;
 Best Local Similarity 47.6%; Pred. No. 5.9e-92;
 Matches 258; Conservative 77; Mismatches 179; Indels 28; Gaps 14;
 QY 73 DWHDKEAIRDAROVNGSGRPYMTDAERYD---QAYRENGFIYVSDKISLNSRLP 128
 DB 27 DYHNYEQIONDINRVPGSGKRPATISPEARSELRKEIYYKNGFVALLSDISIRRSJA 86
 QY 129 DIRHPNCSKRYLETLPNTSIIIPFHEGWSLLRTVHSVLRSPPELVAEIVLVDDFSD 188
 DB 87 DLRHPSCKLSYRSHLPASVAVPYEHEHWSLLRTVSVLNRSPPELVAEIVLVDDGST 146

QY 189 REHLKPELDYML-PSVRLRTKKEGLIRTMLGASVATGDPVITFLDSHCANVNL 247
 DB 147 KEFLHNLDEYVKNLPRKVLVROPERTGLIKARLAKASGVILFLDSHTAGVNL 206
 QY 248 PPLIDRIARRKRTIVCEMIDVIDHDPRFRETQAGDANRGAFDWEMYKRIPIPELOKAD 307
 DB 207 PPLLEPIAENPKTCVCPILIDVIDDQTFDVAHQ--DEGGRLFDWTFYKRVIKNE-DRIS 264
 QY 308 PSDPFEPPVMAAGGFAVADRKKFMELGSDYDRLVWGEQVEISKVMCGGRMEDICSR 367
 DB 265 PTEPFPPVMAAGGFAIGADFPWELGSDYDRLVWGEQVEISKVMCGGRMEDICSR 324
 QY 368 VGHYKRVYKPVKAVGVS-LARNLRKVAVWMDVAYATVQRRRE-VRHLIAGDVAQKK 425
 DB 325 FGHYKRVYKPVKAVGVS-LARNLRKVAVWMDVAYATVQRRRE-VRHLIAGDVAQKK 384
 QY 426 LRSLNCKSPKFWMTKIAMDLPKFYVPVEPPAAAGBIRNVG-TGLCADTKGALGSPLR 484
 DB 385 IREKLMCKPKFMFLQEVAPRIELVPPVEPPVYASGSIQVADSSLCIDTMQRGRGPIG 444
 QY 485 LBGVRRGRGAANNVQVFTFTWEDIRPGDPQTKKF---CPDAISHT--SPVTLVDC 539
 DB 445 LYPGNSNLIIEPT-NHNOYFVHSMHRDI-----QH--KYGGCQPDVPSKQSPVITFTCH 496
 QY 540 SMKGNOLMKYRKDKTLVHPVSGSCMD-----CSGSDHRIFMWTCNPSSLTQOMLFHNTST 595
 DB 497 MHQGNQFPQY-DHKTQOIKRNGVCTISDSPAKECHRNINYSKQLPQTSIIVPFDEHWT 555
 QY 596 VL 597
 DB 556 LL 557

RESULT 12

GLT4 DROME STANDARD; PRT; 659 AA.

AC 081A2; 081011.
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE N-acetylglucosaminyltransferase 4 (EC 2.4.1.-) (Protein-UDP
 acetylglucosaminyltransferase 4) (UDP-GalNAc:polypeptide N-
 acetylglucosaminyltransferase 4) (pp-GalNAc 4).
 CN Name:pgant4; ORFNames:CG31956;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., ENZYME ACTIVITY, TISSUE SPECIFICITY, AND
 RP DEVELOPMENTAL STAGE.
 RC STRAIN=Canton-S; TISSUE=Embryo;
 RX MEDLINE=22841110; PubMed=12829714; DOI=10.1074/jbc.M3036200;
 RA Ten Hagen K.G., Tran T.T., Gecken T.A., Stein D.S., Zhang Z.;
 RT "Functional characterization and expression analysis of members of the
 RT UDP-GalNAc:polypeptide N-acetylglucosaminyltransferase family from
 RT Drosophila melanogaster."
 RL J. Biol. Chem. 278:35039-35048 (2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.P.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champ M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Baau A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Goadec A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
 RA Isailin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Koditz C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeok P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Morkulov G., Mikhina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirbaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review."
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE OF 16-659 FROM N.A.
 RX PubMed=11925450; DOI=10.1074/jbc.M202664200;
 RA Schwenter T., Bennett E.P., Flores C., Thacker J., Hollmann M.,
 RA Reis C.A., Behrens J., Mandel U., Kech B., Schaefer M.A.,
 RA Haselmann K., Zubarev R., Roepstorff P., Burchell J.M.,
 RA Taylor-Papadimitriou J., Hollingsworth M.A., Clausen H.;
 RT "Functional conservation of subfamilies of putative UDP-N-
 acetylglucosamine:polypeptide N-acetylglucosaminyltransferases in
 RT Drosophila, Caenorhabditis elegans, and mammals. One subfamily
 RT composed of 1(2)35aa is essential in Drosophila."
 RL J. Biol. Chem. 277:22623-22638(2002).
 CC -1- FUNCTION: Glycopeptide transferase involved in O-linked
 CC oligosaccharide biosynthesis, which catalyzes the transfer of an
 CC N-acetyl-D-glucosamine residue to an already glycosylated
 CC peptide. In contrast to other proteins of the family, it does not
 CC act as a peptide transferase that transfers GalNAc onto serine or
 CC threonine residue on the protein receptor, but instead requires
 CC the prior addition of a GalNAc on a peptide before adding
 CC additional GalNAc moieties. Some peptide transferase activity is
 CC however not excluded, considering that its appropriate peptide
 CC substrate may remain unidentified. Prefers the diglycosylated
 CC Nucleoside-3/13 as substrate.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 CC similarity).
 CC -1- TISSUE SPECIFICITY: Expressed in developing oocytes and egg
 CC chambers.
 CC -1- DEVELOPMENTAL STAGE: Expressed during embryonic, larval, pupal and
 CC adult stages. Weakly expressed during early embryonic stages but

```

CC displays a dramatic increase at 12-24 h of embryonic development.
CC continues to be in adult but displays much lower levels in the
CC female adult as compared with the male.
CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
CC region: the N-terminal domain (domain A, also called G1 motif),
CC which is probably involved in manganese coordination and substrate
CC binding and the C-terminal domain (domain B, also called
CC Gal/GalNAc-T motif), which is probably involved in catalytic
CC reaction and UDP-Gal binding (By similarity).
CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
CC contributes to the glycopeptide specificity (By similarity).
CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
CC subfamily.
CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-rib.ch/announce/
CC or send an email to license@isb-rib.ch).
CC -----
CC EMBL, AY268065; AAQ56701.1; -.
CC EMBL, AE003579; AAN10370.1; -.
CC EMBL, AF24752; AAN75751.1; -.
CC FLYBASE: FBgn0051956; Pgnat4.
CC DR InterPro: IPR001173; Glyco trans_2.
CC DR InterPro: IPR008997; RicinB_like.
CC DR InterPro: IPR000772; Ricin_B_lectin.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC DR Pfam: PF00652; Ricin_B_lectin; 3.
CC DR POSITIVE: PSS0231; RICIN_B_LECTIN; 1.
CC DR Calcium; Glycosyltransferase; Golgi stack; Lectin; Manganese;
CC Signal-anchor; Transferase; Transmembrane.
CC FT DOMAIN 1 25
CC TRANSMEM 1 48
CC FT 1 25
CC FT 2 48
CC FT 3 48
CC FT 4 48
CC FT 5 48
CC FT 6 48
CC FT 7 48
CC FT 8 48
CC FT 9 48
CC FT 10 48
CC FT 11 48
CC FT 12 48
CC FT 13 48
CC FT 14 48
CC FT 15 48
CC FT 16 48
CC FT 17 48
CC FT 18 48
CC FT 19 48
CC FT 20 48
CC FT 21 48
CC FT 22 48
CC FT 23 48
CC FT 24 48
CC FT 25 48
CC FT 26 48
CC FT 27 48
CC FT 28 48
CC FT 29 48
CC FT 30 48
CC FT 31 48
CC FT 32 48
CC FT 33 48
CC FT 34 48
CC FT 35 48
CC FT 36 48
CC FT 37 48
CC FT 38 48
CC FT 39 48
CC FT 40 48
CC FT 41 48
CC FT 42 48
CC FT 43 48
CC FT 44 48
CC FT 45 48
CC FT 46 48
CC FT 47 48
CC FT 48 48
CC FT 49 48
CC FT 50 48
CC FT 51 48
CC FT 52 48
CC FT 53 48
CC FT 54 48
CC FT 55 48
CC FT 56 48
CC FT 57 48
CC FT 58 48
CC FT 59 48
CC FT 60 48
CC FT 61 48
CC FT 62 48
CC FT 63 48
CC FT 64 48
CC FT 65 48
CC FT 66 48
CC FT 67 48
CC FT 68 48
CC FT 69 48
CC FT 70 48
CC FT 71 48
CC FT 72 48
CC FT 73 48
CC FT 74 48
CC FT 75 48
CC FT 76 48
CC FT 77 48
CC FT 78 48
CC FT 79 48
CC FT 80 48
CC FT 81 48
CC FT 82 48
CC FT 83 48
CC FT 84 48
CC FT 85 48
CC FT 86 48
CC FT 87 48
CC FT 88 48
CC FT 89 48
CC FT 90 48
CC FT 91 48
CC FT 92 48
CC FT 93 48
CC FT 94 48
CC FT 95 48
CC FT 96 48
CC FT 97 48
CC FT 98 48
CC FT 99 48
CC FT 100 48
CC FT 101 48
CC FT 102 48
CC FT 103 48
CC FT 104 48
CC FT 105 48
CC FT 106 48
CC FT 107 48
CC FT 108 48
CC FT 109 48
CC FT 110 48
CC FT 111 48
CC FT 112 48
CC FT 113 48
CC FT 114 48
CC FT 115 48
CC FT 116 48
CC FT 117 48
CC FT 118 48
CC FT 119 48
CC FT 120 48
CC FT 121 48
CC FT 122 48
CC FT 123 48
CC FT 124 48
CC FT 125 48
CC FT 126 48
CC FT 127 48
CC FT 128 48
CC FT 129 48
CC FT 130 48
CC FT 131 48
CC FT 132 48
CC FT 133 48
CC FT 134 48
CC FT 135 48
CC FT 136 48
CC FT 137 48
CC FT 138 48
CC FT 139 48
CC FT 140 48
CC FT 141 48
CC FT 142 48
CC FT 143 48
CC FT 144 48
CC FT 145 48
CC FT 146 48
CC FT 147 48
CC FT 148 48
CC FT 149 48
CC FT 150 48
CC FT 151 48
CC FT 152 48
CC FT 153 48
CC FT 154 48
CC FT 155 48
CC FT 156 48
CC FT 157 48
CC FT 158 48
CC FT 159 48
CC FT 160 48
CC FT 161 48
CC FT 162 48
CC FT 163 48
CC FT 164 48
CC FT 165 48
CC FT 166 48
CC FT 167 48
CC FT 168 48
CC FT 169 48
CC FT 170 48
CC FT 171 48
CC FT 172 48
CC FT 173 48
CC FT 174 48
CC FT 175 48
CC FT 176 48
CC FT 177 48
CC FT 178 48
CC FT 179 48
CC FT 180 48
CC FT 181 48
CC FT 182 48
CC FT 183 48
CC FT 184 48
CC FT 185 48
CC FT 186 48
CC FT 187 48
CC FT 188 48
CC FT 189 48
CC FT 190 48
CC FT 191 48
CC FT 192 48
CC FT 193 48
CC FT 194 48
CC FT 195 48
CC FT 196 48
CC FT 197 48
CC FT 198 48
CC FT 199 48
CC FT 200 48
CC FT 201 48
CC FT 202 48
CC FT 203 48
CC FT 204 48
CC FT 205 48
CC FT 206 48
CC FT 207 48
CC FT 208 48
CC FT 209 48
CC FT 210 48
CC FT 211 48
CC FT 212 48
CC FT 213 48
CC FT 214 48
CC FT 215 48
CC FT 216 48
CC FT 217 48
CC FT 218 48
CC FT 219 48
CC FT 220 48
CC FT 221 48
CC FT 222 48
CC FT 223 48
CC FT 224 48
CC FT 225 48
CC FT 226 48
CC FT 227 48
CC FT 228 48
CC FT 229 48
CC FT 230 48
CC FT 231 48
CC FT 232 48
CC FT 233 48
CC FT 234 48
CC FT 235 48
CC FT 236 48
CC FT 237 48
CC FT 238 48
CC FT 239 48
CC FT 240 48
CC FT 241 48
CC FT 242 48
CC FT 243 48
CC FT 244 48
CC FT 245 48
CC FT 246 48
CC FT 247 48
CC FT 248 48
CC FT 249 48
CC FT 250 48
CC FT 251 48
CC FT 252 48
CC FT 253 48
CC FT 254 48
CC FT 255 48
CC FT 256 48
CC FT 257 48
CC FT 258 48
CC FT 259 48
CC FT 260 48
CC FT 261 48
CC FT 262 48
CC FT 263 48
CC FT 264 48
CC FT 265 48
CC FT 266 48
CC FT 267 48
CC FT 268 48
CC FT 269 48
CC FT 270 48
CC FT 271 48
CC FT 272 48
CC FT 273 48
CC FT 274 48
CC FT 275 48
CC FT 276 48
CC FT 277 48
CC FT 278 48
CC FT 279 48
CC FT 280 48
CC FT 281 48
CC FT 282 48
CC FT 283 48
CC FT 284 48
CC FT 285 48
CC FT 286 48
CC FT 287 48
CC FT 288 48
CC FT 289 48
CC FT 290 48
CC FT 291 48
CC FT 292 48
CC FT 293 48
CC FT 294 48
CC FT 295 48
CC FT 296 48
CC FT 297 48
CC FT 298 48
CC FT 299 48
CC FT 300 48
CC FT 301 48
CC FT 302 48
CC FT 303 48
CC FT 304 48
CC FT 305 48
CC FT 306 48
CC FT 307 48
CC FT 308 48
CC FT 309 48
CC FT 310 48
CC FT 311 48
CC FT 312 48
CC FT 313 48
CC FT 314 48
CC FT 315 48
CC FT 316 48
CC FT 317 48
CC FT 318 48
CC FT 319 48
CC FT 320 48
CC FT 321 48
CC FT 322 48
CC FT 323 48
CC FT 324 48
CC FT 325 48
CC FT 326 48
CC FT 327 48
CC FT 328 48
CC FT 329 48
CC FT 330 48
CC FT 331 48
CC FT 332 48
CC FT 333 48
CC FT 334 48
CC FT 335 48
CC FT 336 48
CC FT 337 48
CC FT 338 48
CC FT 339 48
CC FT 340 48
CC FT 341 48
CC FT 342 48
CC FT 343 48
CC FT 344 48
CC FT 345 48
CC FT 346 48
CC FT 347 48
CC FT 348 48
CC FT 349 48
CC FT 350 48
CC FT 351 48
CC FT 352 48
CC FT 353 48
CC FT 354 48
CC FT 355 48
CC FT 356 48
CC FT 357 48
CC FT 358 48
CC FT 359 48
CC FT 360 48
CC FT 361 48
CC FT 362 48
CC FT 363 48
CC FT 364 48
CC FT 365 48
CC FT 366 48
CC FT 367 48
CC FT 368 48
CC FT 369 48
CC FT 370 48
CC FT 371 48
CC FT 372 48
CC FT 373 48
CC FT 374 48
CC FT 375 48
CC FT 376 48
CC FT 377 48
CC FT 378 48
CC FT 379 48
CC FT 380 48
CC FT 381 48
CC FT 382 48
CC FT 383 48
CC FT 384 48
CC FT 385 48
CC FT 386 48
CC FT 387 48
CC FT 388 48
CC FT 389 48
CC FT 390 48
CC FT 391 48
CC FT 392 48
CC FT 393 48
CC FT 394 48
CC FT 395 48
CC FT 396 48
CC FT 397 48
CC FT 398 48
CC FT 399 48
CC FT 400 48
CC FT 401 48
CC FT 402 48
CC FT 403 48
CC FT 404 48
CC FT 405 48
CC FT 406 48
CC FT 407 48
CC FT 408 48
CC FT 409 48
CC FT 410 48
CC FT 411 48
CC FT 412 48
CC FT 413 48
CC FT 414 48
CC FT 415 48
CC FT 416 48
CC FT 417 48
CC FT 418 48
CC FT 419 48
CC FT 420 48
CC FT 421 48
CC FT 422 48
CC FT 423 48
CC FT 424 48
CC FT 425 48
CC FT 426 48
CC FT 427 48
CC FT 428 48
CC FT 429 48
CC FT 430 48
CC FT 431 48
CC FT 432 48
CC FT 433 48
CC FT 434 48
CC FT 435 48
CC FT 436 48
CC FT 437 48
CC FT 438 48
CC FT 439 48
CC FT 440 48
CC FT 441 48
CC FT 442 48
CC FT 443 48
CC FT 444 48
CC FT 445 48
CC FT 446 48
CC FT 447 48
CC FT 448 48
CC FT 449 48
CC FT 450 48
CC FT 451 48
CC FT 452 48
CC FT 453 48
CC FT 454 48
CC FT 455 48
CC FT 456 48
CC FT 457 48
CC FT 458 48
CC FT 459 48
CC FT 460 48
CC FT 461 48
CC FT 462 48
CC FT 463 48
CC FT 464 48
CC FT 465 48
CC FT 466 48
CC FT 467 48
CC FT 468 48
CC FT 469 48
CC FT 470 48
CC FT 471 48
CC FT 472 48
CC FT 473 48
CC FT 474 48
CC FT 475 48
CC FT 476 48
CC FT 477 48
CC FT 478 48
CC FT 479 48
CC FT 480 48
CC FT 481 48
CC FT 482 48
CC FT 483 48
CC FT 484 48
CC FT 485 48
CC FT 486 48
CC FT 487 48
CC FT 488 48
CC FT 489 48
CC FT 490 48
CC FT 491 48
CC FT 492 48
CC FT 493 48
CC FT 494 48
CC FT 495 48
CC FT 496 48
CC FT 497 48
CC FT 498 48
CC FT 499 48
CC FT 500 48
CC FT 501 48
CC FT 502 48
CC FT 503 48
CC FT 504 48
CC FT 505 48
CC FT 506 48
CC FT 507 48
CC FT 508 48
CC FT 509 48
CC FT 510 48
CC FT 511 48
CC FT 512 48
CC FT 513 48
CC FT 514 48
CC FT 515 48
CC FT 516 48
CC FT 517 48
CC FT 518 48
CC FT 519 48
CC FT 520 48
CC FT 521 48
CC FT 522 48
CC FT 523 48
CC FT 524 48
CC FT 525 48
CC FT 526 48
CC FT 527 48
CC FT 528 48
CC FT 529 48
CC FT 530 48
CC FT 531 48
CC FT 532 48
CC FT 533 48
CC FT 534 48
CC FT 535 48
CC FT 536 48
CC FT 537 48
CC FT 538 48
CC FT 539 48
CC FT 540 48
CC FT 541 48
CC FT 542 48
CC FT 543 48
CC FT 544 48
CC FT 545 48
CC FT 546 48
CC FT 547 48
CC FT 548 48
CC FT 549 48
CC FT 550 48
CC FT 551 48
CC FT 552 48
CC FT 553 48
CC FT 554 48
CC FT 555 48
CC FT 556 48
CC FT 557 48
CC FT 558 48
CC FT 559 48
CC FT 560 48
CC FT 561 48
CC FT 562 48
CC FT 563 48
CC FT 564 48
CC FT 565 48
CC FT 566 48
CC FT 567 48
CC FT 568 48
CC FT 569 48
CC FT 570 48
CC FT 571 48
CC FT 572 48
CC FT 573 48
CC FT 574 48
CC FT 575 48
CC FT 576 48
CC FT 577 48
CC FT 578 48
CC FT 579 48
CC FT 580 48
CC FT 581 48
CC FT 582 48
CC FT 583 48
CC FT 584 48
CC FT 585 48
CC FT 586 48
CC FT 587 48
CC FT 588 48
CC FT 589 48
CC FT 590 48
CC FT 591 48
CC FT 592 48
CC FT 593 48
CC FT 594 48
CC FT 595 48
CC FT 596 48
CC FT 597 48
CC FT 598 48
CC FT 599 48
CC FT 600 48
CC FT 601 48
CC FT 602 48
CC FT 603 48
CC FT 604 48
CC FT 605 48
CC FT 606 48
CC FT 607 48
CC FT 608 48
CC FT 609 48
CC FT 610 48
CC FT 611 48
CC FT 612 48
CC FT 613 48
CC FT 614 48
CC FT 615 48
CC FT 616 48
CC FT 617 48
CC FT 618 48
CC FT 619 48
CC FT 620 48
CC FT 621 48
CC FT 622 48
CC FT 623 48
CC FT 624 48
CC FT 625 48
CC FT 626 48
CC FT 627 48
CC FT 628 48
CC FT 629 48
CC FT 630 48
CC FT 631 48
CC FT 632 48
CC FT 633 48
CC FT 634 48
CC FT 635 48
CC FT 636 48
CC FT 637 48
CC FT 638 48
CC FT 639 48
CC FT 640 48
CC FT 641 48
CC FT 642 48
CC FT 643 48
CC FT 644 48
CC FT 645 48
CC FT 646 48
CC FT 647 48
CC FT 648 48
CC FT 649 48
CC FT 650 48
CC FT 651 48
CC FT 652 48
CC FT 653 48
CC FT 654 48
CC FT 655 48
CC FT 656 48
CC FT 657 48
CC FT 658 48
CC FT 659 48
CC FT 660 48
CC FT 661 48
CC FT 662 48
CC FT 663 48
CC FT 664 48
CC FT 665 48
CC FT 666 48
CC FT 667 48
CC FT 668 48
CC FT 669 48
CC FT 670 48
CC FT 671 48
CC FT 672 48
CC FT 673 48
CC FT 674 48
CC FT 675 48
CC FT 676 48
CC FT 677 48
CC FT 678 48
CC FT 679 48
CC FT 680 48
CC FT 681 48
CC FT 682 48
CC FT 683 48
CC FT 684 48
CC FT 685 48
CC FT 686 48
CC FT 687 48
CC FT 688 48
CC FT 689 48
CC FT 690 48
CC FT 691 48
CC FT 692 48
CC FT 693 48
CC FT 694 48
CC FT 695 48
CC FT 696 48
CC FT 697 48
CC FT 698 48
CC FT 699 48
CC FT 700 48
CC FT 701 48
CC FT 702 48
CC FT 703 48
CC FT 704 48
CC FT 705 48
CC FT 706 48
CC FT 707 48
CC FT 708 48
CC FT 709 48
CC FT 710 48
CC FT 711 48
CC FT 712 48
CC FT 713 48
CC FT 714 48
CC FT 715 48
CC FT 716 48
CC FT 717 48
CC FT 718 48
CC FT 719 48
CC FT 720 48
CC FT 721 48
CC FT 722 48
CC FT 723 48
CC FT 724 48
CC FT 725 48
CC FT 726 48
CC FT 727 48
CC FT 728 48
CC FT 729 48
CC FT 730 48
CC FT 731 48
CC FT 732 48
CC FT 733 48
CC FT 734 48
CC FT 735 48
CC FT 736 48
CC FT 737 48
CC FT 738 48
CC FT 739 48
CC FT 740 48
CC FT 741 48
CC FT 742 48
CC FT 743 48
CC FT 744 48
CC FT 745 48
CC FT 746 48
CC FT 747 48
CC FT 748 48
CC FT 749 48
CC FT 750 48
CC FT 751 48
CC FT 752 48
CC FT 753 48
CC FT 754 48
CC FT 755 48
CC FT 756 48
CC FT 757 48
CC FT 758 48
CC FT 759 48
CC FT 760 48
CC FT 761 48
CC FT 762 48
CC FT 763 48
CC FT 764 48
CC FT 765 48
CC FT 766 48
CC FT 767 48
CC FT 768 48
CC FT 769 48
CC FT 770 48
CC FT 771 48
CC FT 772 48
CC FT 773 48
CC FT 774 48
CC FT 775 48
CC FT 776 48
CC FT 777 48
CC FT 778 48
CC FT 779 48
CC FT 780 48
CC FT 781 48
CC FT 782 48
CC FT 783 48
CC FT 784 48
CC FT 785 48
CC FT 786 48
CC FT 787 48
CC FT 788 48
CC FT 789 48
CC FT 790 48
CC FT 791 48
CC FT 792 48
CC FT 793 48
CC FT 794 48
CC FT 795 48
CC FT 796 48
CC FT 797 48
CC FT 798 48
CC FT 799 48
CC FT 800 48
CC FT 801 48
CC FT 802 48
CC FT 803 48
CC FT 804 48
CC FT 805 48
CC FT 806 48
CC FT 807 48
CC FT 808 48
CC FT 809 48
CC FT 810 48
CC FT 811 48
CC FT 812 48
CC FT 813 48
CC FT 814 48
CC FT 815 48
CC FT 816 48
CC FT 817 48
CC FT 818 48
CC FT 819 48
CC FT 820 48
CC FT 821 48
CC FT 822 48
CC FT 823 48
CC FT 824 48
CC FT 825 48
CC FT 826 48
CC FT 827 48
CC FT 828 48
CC FT 829 48
CC FT 830 48
CC FT 831 48
CC FT 832 48
CC FT 833 48
CC FT 834 48
CC FT 835 48
CC FT 836 48
CC FT 837 48
CC FT 838 48
CC FT 839 48
CC FT 840 48
CC FT 841 48
CC FT 842 48
CC FT 843 48
CC FT 844 48
CC FT 845 48
CC FT 846 48
CC FT 847 48
CC FT 848 48
CC FT 849 48
CC FT 850 48
CC FT 851 48
CC FT 852 48
CC FT 853 48
CC FT 854 48
CC FT 855 48
CC FT 856 48
CC FT 857 48
CC FT 858 48
CC FT 859 48
CC FT 860 48
CC FT 861 48
CC FT 862 48
CC FT 863 48
CC FT 864 48
CC FT 865 48
CC FT 866 48
CC FT 867 48
CC FT 868 48
CC FT 869 48
CC FT 870 48
CC FT 871 48
CC FT 872 48
CC FT 873 48
CC FT 874 48
CC FT 875 48
CC FT 876 48
CC FT 877 48
CC FT 878 48
CC FT 879 48
CC FT 880 48
CC FT 881 48
CC FT 882 48
CC FT 883 48
CC FT 884 48
CC FT 885 48
CC FT 886 48
CC FT 887 48
CC FT 888 48
CC FT 889 48
CC FT 890 48
CC FT 891 48
CC FT 892 48
CC FT 893 48
CC FT 894 48
CC FT 895 48
CC FT 896 48
CC FT 897 48
CC FT 898 48
CC FT 899 48
CC FT 900 48
CC FT 901 48
CC FT 902 48
CC FT 903 48
CC FT 904 48
CC FT 905 48
CC FT 906 48
CC FT 907 48
CC FT 908 48
CC FT 909 48
CC FT 910 48
CC FT 911 48
CC FT 912 48
CC FT 913 48
CC FT 914 48
CC FT 915 48
CC FT 916 48
CC FT 917 48
CC FT 918 48
CC FT 919 48
CC FT 920 48
CC FT 921 48
CC FT 922 48
CC FT 923 48
CC FT 924 48
CC FT 925 48
CC FT 926 48
CC FT 927 48
CC FT 928 48
CC FT 929 48
CC FT 930 48
CC FT 931 48
CC FT 932 48
CC FT 933 48
CC FT 934 48
CC FT 935 48
CC FT 936 48
CC FT 937 48
CC FT 938 48
CC FT 939 48
CC FT 940 48
CC FT 941 48
CC FT 942 48
CC FT 943 48
CC FT 944 48
CC FT 945 48
CC FT 946 48
CC FT 947 48
CC FT 948 48
CC FT 949 48
CC FT 950 48
CC FT 951 48
CC FT 952 48
CC FT 953 48
CC FT 954 48
CC FT 955 48
CC FT 956 48
CC FT 957 48
CC FT 958 48
CC FT 959 48
CC FT 960 48
CC FT 961 48
CC FT 962 48
CC FT 963 48
CC FT 964 48
CC FT 965 48
CC FT 966 48
CC FT 967 48
CC FT 968 48
CC FT 969 48
CC FT 970 48
CC FT 971 48
CC FT 972 48
CC FT 973 48
CC FT 974 48
CC FT 975 48
CC FT 976 48
CC FT 977 48
CC FT 978 48
CC FT 979 48
CC FT 980 48
CC FT 981 48
CC FT 982 48
CC FT 983 48
CC FT 984 48
CC FT 985 48
CC FT 986 48
CC FT 987 48
CC FT 988 48
CC FT 989 48
CC FT 990 48
CC FT 991 48
CC FT 992 48
CC FT 993 48
CC FT 994 48
CC FT 995 48
CC FT 996 48
CC FT 997 48
CC FT 998 48
CC FT 999 48
CC FT 1000 48
CC FT 1001 48
CC FT 1002 48
CC FT 1003 48
CC FT 1004 48
CC FT 1005 48
CC FT 1006 48
CC FT 1007 48
CC FT 1008 48
CC FT 1009 48
CC FT 1010 48
CC FT 1011 48
CC FT 1012 48
CC FT 1013 48
CC FT 1014 48
CC FT 1015 48
CC FT 1016 48
CC FT 1017 48
CC FT 1018 48
CC FT 1019 48
CC FT 1020 48
CC FT 1021 48
CC FT 1022 48
CC FT 1023 48
CC FT 1024 48
CC FT 1025 48
CC FT 1026 48
CC FT 1027 48
CC FT 1028 48
CC FT 1029 48
CC FT 1030 48
CC FT 1031 48
CC FT 1032 48
CC FT 1033 48
CC FT 1034 48
CC FT 1035 48
CC FT 1036 48
CC FT 1037 48
CC FT 1038 48
CC FT 1039 48
CC FT 1040 48
CC FT 1041 48
CC FT 1042 48
CC FT 1043 48
CC FT 1044 48
CC FT 1045 48
CC FT 1046 48
CC FT 1047 48
CC FT 1048 48
CC FT 1049 48
CC FT 1050 48
CC FT 1051 48
CC FT 1052 48
CC FT 1053 48
CC FT 1054 48
CC FT 1055 48
CC FT 1056 48
CC FT 1057 48
CC FT 1058 48
CC FT 1059 48
CC FT 1060 48
CC FT 1061 48
CC FT 1062 48
CC FT 1063 48
CC FT 1064 48
CC FT 1065 48
CC FT 1066 48
CC FT 1067 48
CC FT 1068 48
CC FT 1069 48
CC FT 1070 48
CC FT 1071 48
CC FT 1072 48
CC FT 1073 48
CC FT 1074 48
CC FT 1075 48
CC FT 1076 48
CC FT 1077 48
CC FT 1078 48
CC FT 1079 48
CC FT 1080 48
CC FT 1081 48
CC FT 1082 48
CC FT 1083 48
CC FT 1084 48
CC FT 1085 48
CC FT 1086 48
CC FT 1087 48
CC FT 1088 48
CC FT 1089 48
CC FT 1090 48
CC FT 1091 48
CC FT 1092 48
CC FT 1093 48
CC FT 1094 48
CC FT 1095 48
CC FT 1096 48
CC FT 1097 48
CC FT 1098 48
CC FT 1099 48
CC FT 1100 48
CC FT 1101 48
CC FT 1102 48
CC FT 1103 48
CC FT 1104 48
CC FT 1105 48
CC FT 1106 48
CC FT 1107 48
CC FT 1108 48
CC FT 1109 48
CC FT 1110 48
CC FT 1111 48
CC FT 1112 48
CC FT 1113 48
CC FT 1114 48
CC FT 1115 48
CC FT 1116 48
CC FT 1117 48
CC FT 1118 48
CC FT 1119 48
CC FT 1120 48
CC FT 1121 48
CC FT 1122 48
CC FT 1123 48
CC FT 1124 48
CC FT 1125 48
CC FT 1126 48
CC FT 1127 48
CC FT 1128 48
CC FT 1129 48
CC FT 1130 48
CC FT 1131 48
CC FT 1132 48
CC FT 1133 48
CC FT 1134 48
CC FT 1135 48
CC FT 1136 48
CC FT 1137 48
CC FT 1138 48
CC FT 1139 48
CC FT 1140 48
CC FT 1141 48
CC FT 1142 48
CC FT 1143 48
CC FT 1144 48
CC FT 1145 48
CC FT 1146 48
CC FT 1147 48
CC FT 1148 48
CC FT 1149 48
CC FT 1150 48
CC FT 1151 48
CC FT 1152 48
CC FT
```

Db 200 IFNEHNTLLRSYIVNRPPELLKQIVLVDDGSEMDVLKQPLDDYVQGHFPHLTVI 259
 Qy 210 RTKRGGLRTMGLASVATGVITPLDSCGANWMLPPLRLRIARNTKIVCPMDIV 269
 Db 260 RNBEROGLIGARIGAKVAVAGVWVFDSDHIEVYVWMLPPLIEPIALNPISICPMDVI 319
 Qy 270 DHDDFFETAGAGMAGAFPMWYVYKRIPIPELQKADPDDPESPMPAGGLFVAVRKM 329
 Db 320 SHEDFFSGNKGARGCFMFKYKQPLVPLP-DALDKMPKPSRPMGGLPAINLDF 378
 Qy 330 WEIGYDPCGIEWGEGEYELSPKVMGGMEDIPCSRGIYR-KYVPYVAGVS-LA 387
 Db 379 WDLGVDQDQIDWGEYELSPKIMWGGLLVPCSRVAHITRGPKPRGNHFA 438
 Qy 388 RLTKRAVWMDYAEIYQRPB-YHLSAGDVAVOKLRSSLSLCKSPFMTKIALMDL 446
 Db 439 KHKRAVAVWMDYAEIYQRPB-YHLSAGDVAVOKLRSSLSLCKSPFMTKIALMDL 498
 Qy 447 PKFYPPVPPAAMGELRNVTGTL-CADTHGALGSPLEEGCVRGGEAAMNNQVET 505
 Db 499 LKPEPPVPPVYAGIIVQVAVNPVCLDNGKSTEEAVGMSCADNTHPQPN--QFEL 556
 Qy 506 TWREDIRPDPQHTKPF---CPDAISHTSP---VTLYDCHSMKGNOLMKY-RKDKTLVH 557
 Db 557 SIFRDLR-----MKGFDVCLDV--HEGPNATVWVMSCHSGGNGQFWYDQOTQRLVH 608
 Qy 558 PVSGS-CMD--CSBDRHIFMNTCPSSLTQOQLFEHTNSTVLEKF 600
 Db 609 GENNKRCLEGFVENGIAKVVANSCENDRQWMEFGVHNTMLDTF 654
 RESULT 14
 GLT9_DROME STANDARD; PRT; 650 AA.
 ID GLT9_DROME STANDARD; PRT; 650 AA.
 AC QMRG9; Q9V770;
 DT 01-OCT-2004 (Rel. 45, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Putative polypeptide N-acetylglucosaminyltransferase 9 (EC 2.4.1.41)
 DE (Protein-UDP acetylglucosaminyltransferase 9) (UDP-
 GalNAc:polypeptide N-acetylglucosaminyltransferase 9) (pp-GalNAc
 9).
 CN Name=pgant9; ORFNames=CG30463; fly).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Ranaivosoa P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 Ananides R.G., Scherer S.E., Holt R.A., Gocayne J.D.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.W., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 Beeson K.Y., Berno P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 Burtis K.C., Busham D.A., Butler H., Cadenat L.B., Davies P.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de la Paoloz B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 Foerster A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
 Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Kamil M., Kalish F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 Merklov G., Milshina N.V., Moberly C., Morris J., Mosher A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spralling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,
 Wang Z.-Y., Waasman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 "The genome sequence of *Drosophila melanogaster*,"
 Science 287:2185-2195 (2000).
 [2]
 RP REVISIONS.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 Hradecky P., Huang Y., Kaminck J.S., Milburn G.H., Prochuk S.E.,
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktoglu L., Berman B.P.,
 Bettencourt B.R., Celiker S.E., de Grey A.D.N.J., Drysdale R.A.,
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 Lewis S.B.,
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 systematic review,"
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
 George R.A., Garin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 Rubin G.M., Celiker S.E.,
 RA "A *Drosophila* full-length cDNA resource,"
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).
 RL CC
 CC -1- FUNCTION: May catalyze the initial reaction in O-linked
 galactosamine biosynthesis, the transfer of an N-acetyl-D-
 galactosamine residue to a serine or threonine residue on the
 protein receptor (By similarity).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-galactosamine + polypeptide =
 UDP + N-acetyl-D-galactosaminyl-polypeptide.
 CC -1- COFACTOR: Manganese and calcium (By similarity).
 CC -1- PATHWAY: Glycosylation.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By
 similarity).
 CC -1- DOMAIN: There are two conserved domains in the glycosyltransferase
 region: the N-terminal domain (domain A, also called G1 motif),
 which is probably involved in manganese coordination and substrate
 binding and the C-terminal domain (domain B, also called
 Gal/GalNAc-T motif), which is probably involved in catalytic
 reaction and UDP-gal binding (By similarity).
 CC -1- DOMAIN: The ricin B-type lectin domain binds to GalNAc and
 contributes to the glycopeptide specificity (By similarity).
 CC -1- SIMILARITY: Belongs to the glycosyltransferase family 2. GalNAc-T
 subfamily.
 CC -1- SIMILARITY: Contains 1 ricin B-type lectin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; AB003806; AAF57964.2; -
 CC EMBL; AY121661; AAF57964.1; -
 CC HSPF; P26514; 1KRM.
 DR FlyBase; FBgn0050463; CG30463.

Qy 507 WRSDIRRGDPQHTKKFCFEDAISHTSPVTLYDCHSMKGNOLWKY-RKDKTLYHPVSGSCMD 565
Db 558 KRHQIMSDN-----CLDASNALGPVNLVRCBGMGNGNEMIIDDEKTIKAVNSGNCILT 611
Qy 566 -CSES DHRI-FMNTCNPPSLTQOWLFE 590
Db 612 RASEDDPSTPLLRPCNYSB-GQOWLMQ 637

Search completed: December 20, 2004, 14:12:57
Job time : 200 secs

This Page Blank (uspto)